NITAZOXANIDE AND THE THIAZOLIDES: INVESTIGATIONS INTO DRUG MEDIATED CELLULAR RESPONSES AND IDENTIFICATION OF NOVEL DRUG INTERACTORS

Thesis submitted in accordance with the requirements of

the University of Liverpool for the degree of Doctor in

Philosophy

by

Andrew Thomas Quaile

November 2009

Acknowledgements

I would like to thank my supervisor Prof. Jonathan Wastling who has been a great help during the last 4 years.

I would also like to extend special thanks to Dr Sanya Sanderson who has dedicated time and effort far beyond the call of duty, and has taught and helped me enormously during the last 4 years and I would like to thank her for everything she has done to make this work possible.

Thanks also go to Mr Dong Xia, Dr Morag Nelson, Miss Rebecca Norton, Dr Nadine Randle and the other members of the Infection Biology group past and present who have been both friends and colleagues and whom I hope will continue to be for many years to come.

This project would not have been possible without the invaluable collaborations which have formed the basis of many parts of this study. Special thanks go to Dr Andrew Stachulski, Dr Mazhar Iqbal and the rest of the Chemical Synthesis Group (Dept. of Chemistry) and also Drs Neil Berry and Prabha Jayapal of Computational Chemistry (Dept. of Chemistry).

Thanks also to our collaborators at the University of Bern, Prof. Norbert Müller, Dr Andrew Hemphill and Dr Joachim Müller. Although I would also like to add a special thankyou to Norbert Müller who was an excellent host and supervised my time in Switzerland and was also not too angry when I accidentally stole his skis half way up a very large mountain.

Thanks also to Dr Jeffrey Glenn (Stanford) and his group for the donation of RP7 replicon cells.

Dedicated to my fiancé Jemma Egan

Authors Declaration

The work presented in this thesis was performed solely by the author except where the assistance of others has been acknowledged.

Andrew Quaile, November 2009

Abstract

Nitazoxanide (NTZ) has been shown to be effective against a wide range of pathogens including anaerobic bacteria, protozoan parasites and several viruses. Previously postulated mechanisms of action cannot satisfactorily explain the wide range of activity of this class of compounds, leading to the hypothesis that there may be multiple mechanisms of action.

The aim of this study was to profile gross biological changes in response to, and interactions with, the thiazolides that might relate to the thiazolides mechanism of action and thus obtain data that might be useful in determining new or more detailed information about this class of compounds broad spectrum of activity.

In this study the cellular responses to nitazoxanide and the other thiazolides against the intestinal parasite *Giardia lamblia* and against hepatitis C virus (HCV) infected cells were investigated. A NTZ resistant strain of *G. lamblia* was compared to the naturally occurring wild-type strain, initially using 2D difference gel electrophoresis (DIGE) and quantitative real time PCR, and then also using microarray data. Analysis of differentially expressed proteins and genes indicated that regulation of the cell cycle, mitosis, and chromosomal organisation is important to the resistant strain. Proteomic experiments in particular indicated that protein metabolism, translational elongation and protein folding were also modulated in the resistant strain. Several changes in cytoskeletal proteins were also detected.

DIGE was also used to investigate the proteomic response of Huh7 cells expressing a sub-genomic replicon of HCV to the thiazolides. Expression changes in several heat shock proteins, particularly BiP, a protein responsible

for coordinating the unfolded protein response, indicated that stress responses in the endoplasmic reticulum are important in the mechanism of action against HCV. Modulation of several proteins with critical roles in the correct functioning of HCV replication complexes were also detected upon treatment with thiazolide.

Specific thiazolide-host protein interactions were also sought using tizoxanide affinity chromatography with Huh7 cells. NAD(P)H quinone oxidoreductase 1 (NQO1) was identified as a previously undiscovered binding partner to tizoxanide (TIZ). It was then shown that several thiazolides inhibit NQO1 at physiologically relevant concentrations, and using a combination of *in vitro* enzyme kinetics and computational modelling of the active site, a putative model for binding that resembles the interaction of dicoumarol with the active site of NQO1 is also proposed.

List of Abbreviations

ACN – Acetonitrile

ARE - Antioxidant response element

ATF6 - Activating transcription factor 6

BiP - Binding immunoglobulin protein (also GRP78, HSPA5)

BSA – Bovine serum albumin

C - HCV core protein

CHAPS - 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

ChIP - Chromatin immunoprecipitation

CV - Column Volume

DAG - Directed acyclic graph

DAPA - DNA to protein array

DAVID - Database for Annotation, Visualization and Integrated Discovery

DHRMXXXX – derivative of the specified thiazolide possessing an additional hydroxyl group

DIGE – Difference gel electrophoresis

DMSO - Dimethylsulphoxide

DPI - Dual polarisation interferometry

DTT - Dithiothreitol

E1/E2 - HCV envelope protein 1/2

EAA - Epoxy activated agarose

EIF2α - Elongation initiation factor 2 α

E-IRES - encephalomyocarditis internal ribosome entry site

ENO1 - Enolase 1

ER - Endoplasmin reticulum

ERSE - Endoplasmic reticulum stress response element

ESI - Electrospray ionisation

ExPASy - Expert Protein Analysis System

FA - Formic acid

FAD - flavin adenine dinucleotide

FC - Fold change

FunCatDB - Functional Classification Database

GE - Gel electrophoresis

GO - Gene ontology

GSTP1 - Glutathione-s-transferase pi

GTA-1 - *G. lamblia trophozoite antigen 1*

HBV - Hepatitis B Virus

HCV – Hepatitis C Virus

HMM - Hidden Markov model

HPLC - High performance liquid chromatography

HSP - Heat shock protein

Huh - Human hepatocellular carcinoma cell line

IAA - Iodoacetamide

IC50 - 50 % inhibition concentration

ICAT - Isotope-coded affinity tags

IEF - Isoelectric focussing

IFN - Interferon

IRE - Inositol-requiring 1

IRES - internal ribosome entry site

ISR - Integrated stress response

iTRAQ - isobaric tag for relative and absolute quantitation

KEGG - Kyoto Encyclopedia of Genes and Genomes

Km - Michaelis constant

LC – liquid chromatography

LOWESS - locally weighted scatterplot smoothing

LTQ – Linear ion trap

MBS - Menadione bisulphite

MEN - Menadione

MET - Metronidazole

METr / s - Metronidazole resistant / sensitive

MS - Mass spectrometry

MS/MS - Tandem mass spectrometry

MudPIT - Multi-dimensional protein identification technology

NAD(P)H - Nicotinamide adenine dinucleotide (phosphate)

NAD+ - Reduced nicotinamide adenine dinucleotide

NAPPA - Nucleic acid programmable protein arrays

NL - Non linear

NQO1 - NAD(P)H Quinone oxidoreductase 1

NS - HCV non-structural protein (as in NS5A)

NTZ – Nitazoxanide

NTZr / s - Nitazoxanide resistant / sensitive

OD - Optical density

ORF - Open reading frame

P21.1 - Protein 21.1

p53 - Tumour suppressor protein 53

PAGE - Polyacrylamide gel electrophoresis

PCR - Polymerase chain reaction

PDB - Protein Data Bank

PDI - Protein disulphide isomerase

PERK - PKR-like endoplasmic reticulum kinase

PFOR - Pyruvate ferredox oxidoreductase

PKR - Protein kinase R

PTM - Post-translational modification

qRT-PCR – Quantitative real-time polymerase chain reaction

RC - Replication complex

RP - Reversed phase

RP7 - Huh7 cell line transformed with subgenomic replicon of HCV

RT - Reverse transcriptase

SAGE - Serial analysis of gene expression

SDS - Sodium dodecyl sulphate

SILAC - Stable isotope labelling with amino acids in cell culture

SMC - Structural maintenance of chromosomes

STRING - Search Tool for the Retrieval of Interacting Genes/Proteins

SVR - Sustained virologic response

TAP - Tandem affinity purification

TCA - Trichloroacetic acid

TFA - Trifluoroacetic acid

TIZ – Tizoxanide

TKT - Transketolase

TPP - Thiamine pyrophosphate

TRP - Tryptophan

TYR - Tyrosine

U - Enzyme units

UPR - Unfolded protein response

VSP - Variable surface protein

WTSS - Whole transcriptome shotgun sequencing

XBP1 - X-box binding protein 1

ΔG_b - Gibbs free energy of binding

Table of Contents

Chapter 1

General Introduction

This introductory chapter provides information and context that is of particular importance to the understanding of the subsequent chapters of this thesis. This includes a review of past and current literature and also provides general information about the organisms, diseases, cell types used in the study and also a General Materials and Methods section that details methodologies that are common to more than one chapter or section of this thesis.

This study is centered around a small molecule, anti-infective drug, nitazoxanide (NTZ), that was discovered in 1980 by Euzeby *et al.* [1] (Figure 1.1) as an effective anti-helminthic in cats, dogs and sheep, yet NTZ remained a relatively unknown compound for quite some time. Current records indicate that prior to 1996 there were only three published references it NTZ, all published in the years immediately after its discovery. In the decade that followed, published research into the activity of NTZ increased markedly (Figure 1.2), largely due to the discovery of its efficacy against several parasites [2], a wide range of anaerobic and microaerophillic bacteria [3, 4] and against *Cryptosporidium parvum* and *G. lamblia* [5-7]. However, aside from the routine testing of NTZ for efficacy against new organisms, there was no indication of how NTZ might elicit its effects against such an unusually broad range of pathogens. Another major factor contributing to the increase in research into the thiazolides, particularly during 2008, was the discovery of

their anti-viral properties. Hepatitis B virus (HBV) and Hepatitis C virus (HCV) [8-18] being the most prominent in the literature, however rotavirus, norovirus and others are also mentioned [8, 19-22]. This discovery is of major clinical importance, and has opened new possibilities for treatment of some debilitating and/or deadly infections. Moreover, the failure of NTZs proposed mechanisms of action against non-viral pathogens to adequately explain its anti-viral activity has highlighted a significant gap in our knowledge. Activity against such disparate pathogens is also one of several indications that multiple mechanisms of action are probable.

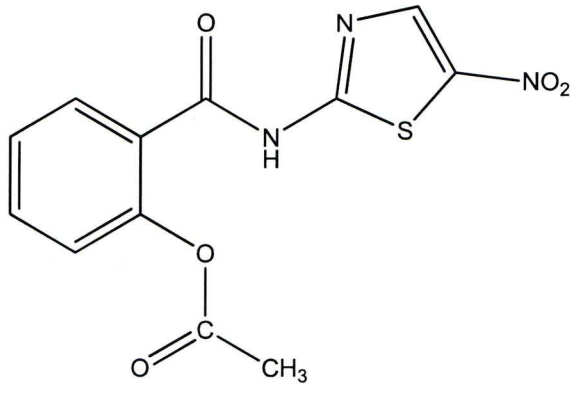


Figure 1.1. The structure of Nitazoxanide. Full chemical name 2-acetyloxy-N-(5-nitro-2-thiazolyl) benzamide.

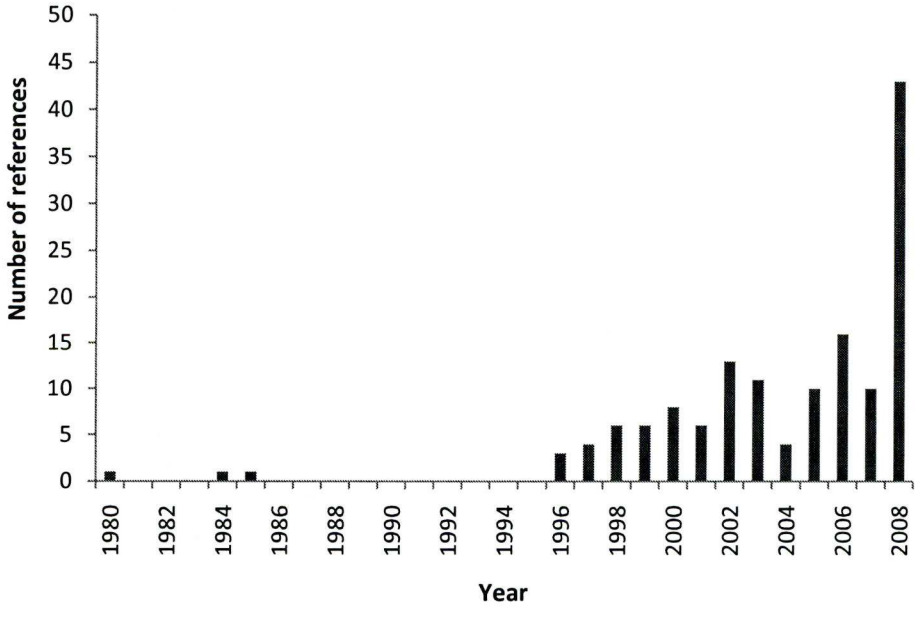


Figure 1.2. Number of published references to Nitazoxanide per year – data retrieved from Web of Science SCI (ISI)

Drugs relating to this study

There are primarily two drugs that relate to the present study, nitazoxanide and metronidazole

Nitazoxanide (NTZ)

The primary compound in this study, NTZ belongs to a class of drugs known as the thiazolides. The core molecule consists of a thiazole moiety and phenyl ring bound by a central peptide bond. Functional groups attached to one or both of these ring structures affect the activity, specificity and efficacy of the various drugs, and also result in different names for the drugs. No standardised naming system for the drugs has been developed, although TIZ, a deacetylated form of NTZ is accepted to be referred to as such, as is the glucuronidated form of TIZ, TIZ-glucuronide. Further derivations of the thiazolides have generally acquired numerical names prefixed by 'RM' that relate to the various 'series' that were synthesised by Liverpool's Dept of Chemistry on behalf of Romark Laboratories. These names and their structural arrangement can be found in the chapters in which they are used.

The mechanism of action of nitazoxanide

For any drug, the mechanism of action is the specific biochemical means by which that drug effects its activity. Understanding the mechanism of action of new and novel drugs is of critical importance to establishing their immediate and long-term safety, improving their efficacy, assessing the potential for the

development of resistance and even in simply developing our understanding of biology as a whole.

In vivo interactions with the thiazolides appear to be complex. Differing functional group modifications have different effects on the various activities of the thiazolides. For example, Muller *et al* [23] found that carboxylation or methylation of position 3 of the benzene ring, or methylation of position 5, diminished or abrogated the activity of the thiazolides against *G. lamblia* parasites. The same effect was observed upon substitution of the nitro group with a bromine residue but bromination does not appear to affect activities against *Neospora caninum* or *Cryptosporidium parvum* (unless also methylated in position 3 of the benzene ring). This has led to the notion that the thiazolides may possess two functionally distinct active centres, at the benzene and nitrothiazole rings and also suggests different targets within the different parasites. However, recent research by Hoffman *et al.* [24] demonstrated a probable mechanism of action of NTZ in selected anaerobic bacteria and *Campylobacter jejuni* via inhibition of pyruvate ferredoxin oxidoreductase (PFOR) through targeting of its activated thiamine pyrophosphate (TPP) cofactor. This proposed mechanism is in contrast to the mechanism of metronidazole which undergoes nitroreduction by PFOR to become activated. This result was particularly interesting, since preventing the interaction with an 'activated' cofactor required for the correct functioning of several enzymes is likely to require more numerous and extensive genetic changes than, for example, a single base pair substitution resulting in reduced affinity between enzyme and inhibitor.

Whilst the discovery of a target in the anaerobic bacteria and parasites tested in their study is of great significance, it remains unclear how this mechanism could account for the efficacy against other organisms, particularly viruses which do not posses these enzymes themselves, and thus the effect must either be mediated through a component of the viruses themselves, or through a host factor. Evidence for the targeting of a host factor, was provided by an investigation into inhibition of hepatitis C virus (HCV) replication. Korba *et al* [25] inferred that since phenotypic resistance to NTZ could not be experimentally transferred to a naïve Huh7.5 cell line by transfer of viral RNA, resistance must be conferred by the host. Although not conclusive, it makes a compelling argument, and is a view increasingly shared amongst members of the field. Unpublished research, personally communicated by Romark Laboratories has suggested that tizoxanide (TIZ), the compound believed to be the circulating active metabolite of NTZ, acts against HCV by disrupting the dephosphorylation of EIF2α that occurs during a typical infection. Or to state it conversely, TIZ action allows the phosphorylation of EIF2α, by PERK/PKR to occur normally, leading potentially to a normal induction of the unfolded protein response (UPR) and integrated stress response (ISR). The UPR and ISR are closely related stress responses that are normally employed in response to stress in the endoplasmic reticulum (ER). Although the interplay between ER stress response mechanisms and HCV is not fully characterised, HCV appears to both induce and control and ER stress response to its own advantage. This is discussed further in Chapter 4. With respect to HCV infection, the structure: activity relationship differs from other

microorganisms; it appears that whilst the nitro position is important, it can be substituted with other groups such as a sulphomethyl (SO_2CH_3) that make it up to ten times more effective against hepatitis C virus (HCV) but ineffective against hepatitis B virus (HBV) *in vitro*. Bromination moderately reduced the efficacy against both HCV and HBV and chlorination abrogated entirely the anti-HCV effects [26].

The complexity of the interactions between the thiazolides, pathogens and their hosts is by no means limited to interactions with PFOR and PKR. NTZ also appears to inhibit the protein disulphide isomerase (PDI) of *Neospora caninum* [27], however it is currently unknown how this mechanism might be effected, and the authors hypothesise that this may not be the primary mechanism of action. A novel nitroreductase of *G. lamblia* is also inhibited by several thiazolides but the functional significance of this has not been fully elucidated and its therapeutic efficacy against this parasite is unknown.

In vivo resistance to nitazoxanide

Since NTZ is often used when treatment with metronidazole has failed, data on the potential for pathogenic resistance to NTZ is scarce and published incidences of resistance in any normally susceptible pathogen appear to be limited to laboratory induced resistance in *G. lamblia* [28]. Although resistance to NTZ could be induced in an Huh7.5 HCV replicon system, it is believed that this is mediated via genetic alterations in host factors over several generations of selective pressure and is thus unlikely to be of

widespread concern with regards to the development of resistance to drug treatment in the general population [29].

Metronidazole

The nitroimidazole antibiotic drug metronidazole (MET) is commonly used for the treatment of giardiasis and a mechanism of action is believed to be known. Pyruvate ferredoxin oxidoreductase (PFOR) reduces ferredoxin, which subsequently transfers it's electrons to the nitro group of MET, resulting in a nitro group that is highly reactive and can form covalent adducts with both protein and DNA [30-32]. Activation by PFOR provides a neat explanation for the spectrum of activity against anaerobic bacteria and parasites that all respire at low or absent concentrations of oxygen and use PFOR to convert pyruvate to acetyl coenzyme A during anaerobic glycolysis. This putative mechanism is supported by the study of MET resistant strains of pathogens that are normally susceptible to MET. In these MET resistant strains, resistance is often associated with reduced expression of PFOR [33, 34]. However, several years after these observations, alternative models for MET resistance are still being proposed. For example, *Trichomonas vaginalis,* which is widely believed to be susceptible to MET via the aforementioned mechanism, has had an alternative flavin-based mechanism of action recently proposed [35].

Gardia lamblia

Hosts and prevalence

Giardia spp. is a micro aerophilic protozoan parasite that resides in the duodenum of the digestive tract [36]. It is a common parasite in cats, dogs, and sheep but it is also prevalent in human populations where it is mainly spread through water supplies contaminated with cysts shed from faeces [37]. It can cause a range of gastrointestinal symptoms, though diarrhoea is the primary problem. Infections are generally chronic and self limiting, however immunocompromised individuals who are already particularly susceptible to infection, will likely require treatment [38, 39]. *G. lamblia* exists in a number of clonal populations, referred to as assemblages, with different host specificities. Humans are affected by two of the more zoonotic assemblages, A (including the highly similar but distinct A1 and A2 sub-assemblages) and B [36]. The isolate used in Chapter 3, WB clone 6, is assemblage A1 and is a clone of an isolate acquired from duodenal aspirate of a 30-year-old American male.

Phylogeny of G. lamblia

Since the first recorded description of *G. lamblia* in 1681 by van Leeuwenhoek [40], its taxonomic and phylogenetic classification has been the subject of some debate. Being a flagellated, protozoan eukaryote, assigning a taxonomic classification to *G. lamblia* based on its visible characteristics finds it in located with species such as *Leishmania* spp., *Trypanosoma* spp. and *Trichomonas vaginalis* [41]. *G. lamblia* is further placed in the order

Diplomonadida, and the Family Hexamitidae. More recently however these taxonomic classifications have been somewhat supplanted by modern phlyogenetic methods relying on the genetic sequences of highly conserved cellular elements such as the small subunit ribosomal RNAs, translation elongation factor 2 and ATPase's [42-44]. These classifications have proven enlightening with regards to the evolutionary position of *G. lamblia*, which places the parasite as an early branching eukaryote [36] (Figure 1.3), although this is still hotly debated by some who argue that the absence of mitochondria but the presence of mitochondrial-like genes means that *G. lamblia* diverged after a proposed endosymbiotic event that may have led to the evolution of the mitochondria, and that these mitochondria were lost as an adaptive change to their low oxygen environment [45, 46].

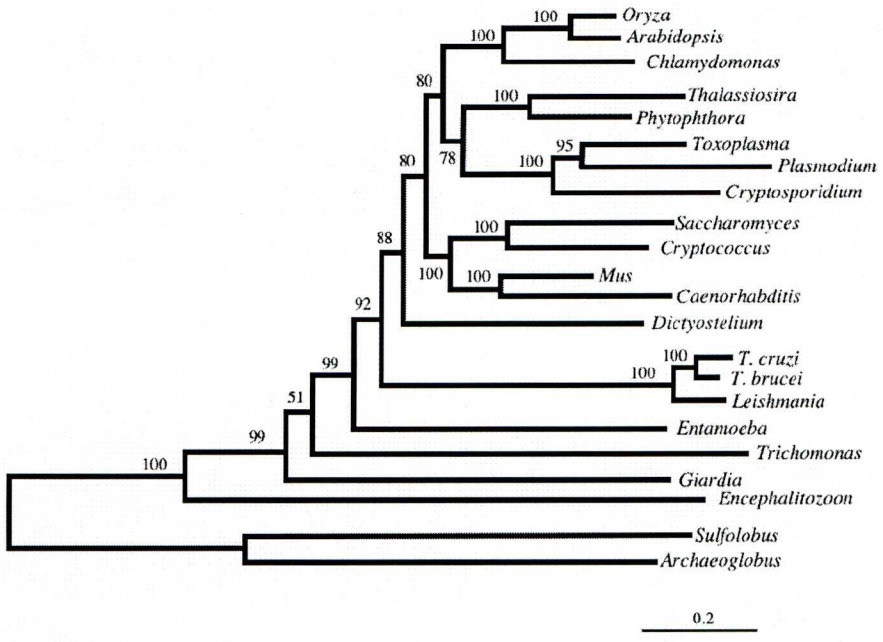


Figure 1.3. Phlogenetic tree comparing 61 ribosomal proteins. Taken from Horizontal lengths are representative of evolutionary change. The scale bar represents the number of amino acid changes per site

Life cycle of G. lamblia

G. lamblia has just two known life cycle stages, the vegetative trophozoite stage and the infective cyst stage. Upon ingestion of cysts and subsequent passage through the stomach, the acidic conditions therein cause excystation of the parasite and subsequently, attachment to the luminal face of the duodenum occurs. It is here that the parasite replicates and exerts any activity that would lead to diarrhoea and malabsorption. *G. lamblia* had long been believed to be asexual, only dividing by mitosis; however there has been significant genetic evidence to suggest that *G. lamblia* can, under the right conditions, undergo sexual meiosis [47, 48] although this has still to be observed directly. Eventually, the growing trophozoites will be exposed to biliary secretions, which will lead some trophozoites to encyst and be passed into the environment along with the faeces, and the infective cycle can recommence. During growth, exchange of nutrients and other factors required for the survival of the parasite and maintenance of the infection occurs via a specialised attachment organelle on the ventral side of the parasite (often referred to simply as the ventral disk), although the precise method of attachment is still the subject of some debate, it appears probably to be a mixture of chemical interactions, and mechanical 'sucking' force, as evidenced by the presence of contractile proteins such as actin, myosin and others [49]. The ventral disk (VD) is made up of a structure of microtubules and microribbons, which are made up from tubulins, and a group of annexin-like proteins referred to as the giardins. Several of these proteins have been localized to these structures [50-53].

Hepatitis C Virus (HCV)

Classification and relationship to other viruses

HCV is an enveloped, positive sense, single stranded RNA virus (Group IV). The sole member of the genus *Hepacivirus,* HCV belongs to the family *Flaviviridae* along with animal pestiviruses (e.g. bovine viral diarrhoea virus) and flaviviruses (e.g. yellow fever, dengue and tick-borne encephalitis viruses) and share similar genome organisation. Unlike the *Pestivirus* and *Flavivirus* which infect a broader variety of species however, HCV causes disease only in humans; though chimpanzees (*Pan troglodytes*) can be experimentally infected [54] and have been historically used as a the only available animal model.

Prevalence

WHO estimates for the worldwide prevalence of HCV infection are approximately 3% (over 200 million based on 2009 population estimates) with higher prevalence in countries in the Far East, Mediterranean countries and certain areas in Africa and Eastern Europe. HCV is normally transmitted via blood-blood contact, commonly through the use of shared injection equipment by drug users or accidental needle sticks however symptoms may arise so long after the infection that it can be difficult to identify the source of infection. Historically however HCV could be transmitted via blood transfusions before screening measures were established to prevent this in the early 1990s.

Genome and viral replication

HCV enters the cell via clathirin mediated endocytosis [55], and uncoats to release the viral RNA into the cytoplasm. An internal ribosome entry site (IRES) in the genome of HCV then forms a complex with the various components required for replication [56]. The HCV genome is approximately 9.6 kilobases in length and is translated to a 3000 amino acid polyprotein that is further cleaved into at least 10 proteins by cellular and viral proteases: the core p21/p19, envelope 1 (E1), E2, p7, nonstructural (NS) 2, NS3, NS4A, NS4B, NS5A, and NS5B. There is also evidence for expression of several frame shifted 'F proteins' (also alternative reading frame proteins - ARFPs) [57, 58] whose functions (if indeed there are any) are still unknown although there is the possibility that they may possess functions that have been previously attributed to the core proteins.

Diseases

Although the immediate (acute) infection with HCV is asymptomatic in ~80 % of cases [59], chronic infections (persisting for over 6 months) develop in 70-90 % of cases [60]. Like the acute infection, chronic infections are usually asymptomatic but become evident when other ailments, including fibrosis and cirrhosis of the liver and subsequently hepatocellular carcinoma arise.

Treatments

The current standard of care for patients harbouring infection with HCV is a combination therapy consisting of pegylated interferon α and ribavirin, but a

sustained viral response (SVR) is only achieved in approximately 50-55 % of cases [61, 62]. Interferon α (IFN-α) acts by inducing the expression of over 300 genes that results in an anti-viral state in the cell [63]. However IFN-α monotherapy has only limited efficacy and typically achieves an SVR in less than 20 % of patients after 6 months of therapy [64]. Ribavirin meanwhile also has limited activity as a monotherapy; however it is able to augment the effects of IFN treatment, raising the combined incidences of SVRs above 50 %. Details of Ribavirin's mechanism of action are still unclear although evidence for several proposed mechanisms exist [65-69].

More recently, NTZ has been identified as a potential treatment for HCV infection for use in combination with PEG-interferon/Ribavirin treatments and is discussed in greater detail in Chapter 4. Other potential new treatments in the pipeline are mainly specific inhibitors of viral enzymes [70] (e.g. the non-structural protein inhibitors telaprevir [71] and boceprevir [72]).

The application of replicon systems

Efficient and reliable replication systems for HCV infection have been historically difficult to find, however the discovery of a sub-genomic replicon system [73] has provided a safe and efficient model for replication that has become widespread tool in HCV research and furthermore has allowed specific functions and interactions of the various components of HCV to be investigated. More details regarding the replicon system can be found in chapter 4.

About 'omic scale approaches

Amongst the more major advances in biology of recent years, the rise of experimental techniques that possess the power to resolve and measure many thousands of biological components (proteins, genes, metabolites etc.) simultaneously has been amongst the most important. Many of these techniques have collectively been suffixed with the term omics; proteomics, genomics, transcriptomics and metabolomics being probably the most common, and has come to be synonymous with the study of a system in its entirety (or a subsystem; e.g. all protein kinases are often referred to as the kinome). Ignoring for a moment the methodological sampling biases of individual techniques, one of the major challenges that is faced by any single 'omics experiment (in particular transcriptomics and proteomics) is that only a portion of the entire picture can be observed using that method. That is to say, investigating the same system with different techniques (e.g. electrophoresis of proteins vs. microarray) will often yield results with only moderate agreement between them [74, 75]. It is for this reason that the central dogma [76, 77] is clearly a vast oversimplification for the purposes of understanding and modelling global cellular changes. Thus it is becoming increasingly apparent that to obtain a more complete and accurate picture of the biology of a system it is important to try to observe that system on multiple and complementary levels at as accurate a resolution as possible. This has led to increased activity in the field of systems biology where a more holistic approach is taken to modelling a biological system.

Proteomics

Proteins are generally attributed to be the "actors" of cellular functions reading from a genetic "script" [78] and as a result, of all of the global cellular profiling techniques, proteomics probably possesses the greatest power to describe the phenotype of an organism. Somewhat ironically however, high throughput protein studies that can distinguish between thousands of protein species are almost entirely reliant on the existence of prior sequence data, and particularly for lesser studied organisms, genomic data and gene prediction algorithms that predict probable proteins. This protein sequence data is searched using the ion masses of peptides derived by mass spectrometry and is a powerful method for the identification of proteins where the protein species is already described or hypothesised. The alternatives to this method of identification are *de novo* sequencing techniques, that although feasible on a smaller scale, are time consuming in the case of Edman degradation or difficult to interpret in the case of tandem mass spectrometry peptide fragmentation data [79-81]. The organisms used in this study both have completely sequenced genomes and so protein species that are of interest can be rapidly identified using peptide and peptide-fragmentation ion mass data.

In the human genome, there are estimated to be approximately 25'000 genes and whilst many of these genes may be inactive depending on the tissue type or environmental conditions, several will possess the potential for alternative splicing before translation into proteins and a great many of these proteins will further be subject to post translational modifications. This means that the techniques employed in any proteomic study should have a very high resolving

power to separate these from one another. Furthermore there are many other challenges in attempting to globally measure proteins, including protein solubility, their wide range of concentrations, protein stability and turnover. The proteomic experiments in this study use 2D gel electrophoresis to separate proteins according to charge, followed by mass, and in combination with sensitive staining can routinely resolve and detect several thousand protein species over greater than 3.5 orders of magnitude of concentration in a single experiment.

Transcriptomics

Transcriptomic technologies such as microarray analysis, whole transcriptome shotgun sequencing (WTSS) such as RNA-seq, and serial analysis of gene expression (SAGE) have established themselves as reliable and informative techniques for profiling the expression of many thousands of genes in a sample. These technologies should be seen as complementary to proteomic methods since they can only describe the level of transcript in a sample and not the protein that it might encode. Microarrays in particular however have certain benefits over other global techniques and lack the statistical problems that are faced by whole proteome experiments. This is thanks to the fact that knowledge of the gene sequences is a prerequisite for synthesis of the chip on which the experiment is performed. As a result, the results can be statistically assessed on the basis of a known population size and replicate experiments are easily combined.

The microarray method has changed little since its inception. Typically short oligonucleotide sequences corresponding to each gene that is to be quantified are synthesised and each unique sequence is attached to the array in a defined area. The total mRNA from a sample is purified and reverse transcribed to cDNA, and samples from control and treatment labelled conditions are labelled with different fluorescent dyes. These samples are hybridised to the oligonucleotide array resulting in a fluorescent signal in the location where the complementary sequence exists. The intensity of the signal corresponds to the concentration of a particular cDNA species. The signals from control and treated conditions are compared and differentially expressed genes are identified [82].

Newer technologies such as RNA-seq do not require prior knowledge of the sequence beforehand, since both sequence and abundance information are obtained at the time of analysis. This has led some to believe that microarrays may soon be replaced by these new technologies [83].

Chapter 2

General Materials and Methods

Any work that attempts to discover drug targets faces a considerable challenge as testing the activity of a drug against every protein species would be an almost insurmountable task. This makes directing the search for targets highly desirable. In this study, we have tried to use several complementary techniques for directing the search for partners interacting with NTZ. One philosophy used for targeting the search for important information that is shared amongst several of the experiments in this study is the use of differential analysis. We have used well established methods of affinity chromatography and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) but in both cases have applied differential techniques that enable us to quickly and effectively eliminate large amounts of 'interfering' data and hone in on the data that will hopefully provide us with useful information. This study also uses several techniques that are shared amongst the different aspects of investigation. Outlined below is an introduction to the theory supporting these shared techniques, some of their strengths and weaknesses and details of the protocols used where it is common between sections of this thesis.

2D-difference gel electrophoresis (2D-DIGE)

The DIGE technique is a combination of 2-Dimensional gel electrophoresis (2D-GE) enhanced by several other experimental and statistical measures

intended to both minimise experimental variation that is common in 2D-GE and reduce the required number of sample replicates for statistical significance. DIGE is undoubtedly a very powerful state of the art technique, but it is not without its inherent systematic limitations. Arguably the greatest limitation results from using a 2D PAGE based separation technique to distinguish between protein species. 2D PAGE based separations are mainly able to resolve proteins that are readily soluble in aqueous solutions. Whereas in 1D SDS-PAGE where sample is added directly to the gel and the use of SDS helps with solubility of some of the more hydrophobic species, IEF does not permit the inclusion of detergents such as these due to the charges imparted on the proteins. Membrane bound or primarily lipid soluble proteins are therefore likely to be underrepresented in 2D PAGE experiments employing IEF as the first dimension separation. Other techniques such a 16-BAC/SDS PAGE [84] have been developed that are particularly suited to the application of resolving membrane proteins, and even 16-BAC/PAGE DIGE experiments but these experiment lack the high resolution of traditional SDS PAGE. Proteins with an extreme pI will also be excluded based on the range of the IEF separation. Co-migration of proteins on the gel can also theoretically lead to ambiguity in assigning an identity or potentially mask significant changes in abundance also [85, 86]. The individual techniques that comprise DIGE used are discussed below.

Isoelectric Focussing

Isoelectric focussing separates proteins on the basis of their net charge at a given pH. Proteins are solubilised and subsequently applied to a gel support (IEF strip) that contains an immobilised pH gradient across its length. When a voltage is applied across this support, proteins move in the direction that will minimise the net charge across the length of the protein. Thus, a protein with a net positive charge at a given pH will move towards the negative electrode until the net charge of the protein becomes zero as a result of the pH. When the proteins reach a point at which the net charge is zero, referred to as their pI, the proteins stop moving. A range of different pH gradients are available that give differing degrees of separation across different pH ranges, these might be useful for example, to obtain a higher resolution separation of a small portion of the proteome (e.g. pH 4-7), however these narrow range strips, will result in various portions of the proteome falling outside their range. Since the experiments here are intended to encompass as broad a sample of the proteome as possible, wide-range strips (pH 3–10 non-linear) are used throughout.

First dimension electrophoresis – Isoelectric focussing method

Immobiline DryStrips (24 cm pH 3 – 10 NL) (GE Healthcare) were rehydrated overnight using 450 µl of Rehydration solution (8 M urea, 2 % (w/v) CHAPS, 0.5 % (v/v) IPG buffer, 20 mM DTT and trace bromophenol). 150 µg of protein (50 µg each from the Cy5, Cy3 and Cy2 labelled samples) were combined as outlined in individual chapters. To maximise mass spectrometry identifications for the *G. lamblia* WT vs NTZr/METr + NTZs/METr DIGE

experiment (Chapter 3), 300 µg of unlabelled protein from undiluted WT4, C42 and C51 samples (100 µg from each) was loaded onto gel number 3. In both the *G. lamblia* WT vs NTZr/METr DIGE and the HCV replicon DIGE experiments 500 µg of protein was loaded onto a separate preparative gel. Finally, samples were made up to a final volume of 100 µl (maximum volume permitted for cup-loading). Samples were applied to the strips using the cup-loading method and run overnight on an Ettan IPGphor II (GE Healthcare) using the protocol outlined in table 2.1

Step	Voltage	Constant/Gradient	Amount	Hours/Volt Hours
1	500	Constant	1	Hours
2	1000	Gradient	5200	Volt hours
3	10000	Gradient	3	Hours
4	10000	Constant	4	Hours

Table 2.1. Isoelectric focusing protocol. Total time was approximately 15 hours. In constant steps, voltage was maintained at the level indicated. In gradient steps voltage was increased to the level specified over the time specified. Duration of individual steps was determined by either a fixed duration of time (Hrs) or by the specified number of volt hours elapsing.

SDS PAGE (including 1D PAGE)

The conditions in the buffers used during SDS-PAGE experiments results in denatured proteins, coated with SDS along their length. This coating is uniform and so proteins are imbued with a negative charge that is proportional to their size. Proteins are transferred to one end of a polyacrylamide gel also containing SDS, in the case of 2D-PAGE, this is directly from an equilibrated IEF strip. A charge is applied across the length of the gel and hence negatively charged proteins are caused to move along the gel towards the anode. As they move, the polyacrylamide acts as a molecular sieve that retards higher molecular weight proteins to a greater extent than those of smaller mass.

2nd dimension electrophoresis – SDS PAGE method

Acrylamide gels (12.5 % (w/v)) were cast simultaneously using the Ettan DALT (GE Healthcare) gel caster and allowed to polymerise overnight. DryStrips were equilibrated for 15 min in equilibration solution (50 mM Tris, 6 M Urea, 30 % (v/v) glycerol, 2 % (w/v) SDS, trace bromophenol blue, pH 8.8) containing first 0.5 % (w/v) dithiotreitol (DTT) then a further 15 min in equilibration solution containing 4.5 % (w/v) iodoacetamide (IAA) at room temperature. Strips were then placed on the top edge of the gels and sealed into place using 0.5 % (w/v) agarose sealing solution also containing bromophenol (trace amount). All gels were run in parallel using the Ettan DALT*six* (GE Healthcare) (8 W per gel, 20 °C, until the bromophenol dye front reached the end of the gel, ~4.5 h).

Cyanine dye labelling

In order to visualise the proteins on the PAGE gel, minimal labelling is performed. Cy2, 3 and 5 are cyanine based fluorescent dyes that are size and charge matched. These are covalently bound to lysine residues at a concentration that causes approximately 3 % of lysine residues within a given protein to become labelled and as a further result of this ratio, it is statistically likely that only a single lysine residue per protein will be labelled thus improving linearity of the resulting fluorescent signal. This technique also has the added advantage that mass spectrometric identifications can be made from analytical gels (as opposed to those with preparative loads) if desired without any modification to the search algorithms as 97 % of protein remains unmodified. The addition of the CyDye does however mean that, particularly at lower molecular weights, the migration distance of the proteins will be retarded. Therefore it is common, if analytical gels are to be used for mass spectrometric identifications that a post stain is used so that the maximal amount of protein can be excised from the gel. Minimal labelling has a sensitivity that is approximately as sensitive as silver staining. A different chemistry that involves the labelling of cysteines at higher concentrations of CyDyes, or 'saturation labelling', is also used due to its high sensitivity (approximately 125 pg of protein according to manufacturers specifications), however the linearity of the stain mentioned above is lost. As the dyes all have been chosen because they have differing excitation and emission wavelengths, multiple samples can be run on each gel, reducing the total number of gels required for a statistically valid result, and further reducing variation. As the

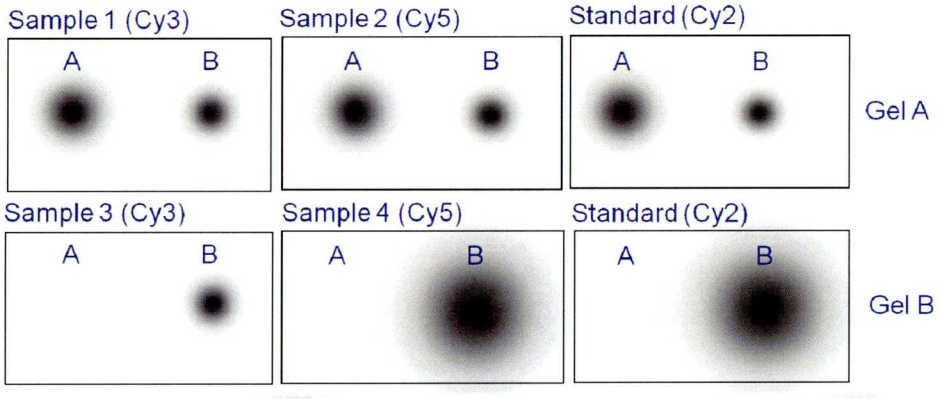
weaker of the fluors, Cy2 is usually used for the internal standard which is discussed below.

CyDye™ minimal labelling method

Labelling was performed according to the CyDye™ minimal labelling protocol (GE Healthcare) using a ratio of 400 pmol dye: 50 µg protein, on ice, in the dark for 30 min, before stopping the reaction with 1 mM lysine on ice for 10 min. Labelled samples were then frozen at -20 °C until required.

Use of a DIGE internal Standard

The accuracy of the DIGE technique is greatly improved by the use of a pooled internal standard. This standard, made up from equal amounts of each biological replicate, ensures that each individual protein is represented on each gel. Thus if a protein appears to be down-regulated to the point of absence in particular sample, reference to the standard can reveal if this is indeed the case, or if the spot is merely missing altogether as can sometimes happen. Inclusion of the standard on every gel in the experiment also aids the automated matching of the spot patterns. The improvement in accuracy that the internal standard provides is further illustrated in Figure 2.1.



Without Std	Sample 1	Sample 2	Sample 3	Sample 4
Spot A	Equal	Equal	Decreased	Decreased
Spot B	Equal	Equal	Equal	Increased

With Std	Sample 1	Sample 2	Sample 3	Sample 4
Spot A	Equal	Equal	Absent	Absent
Spot B	Equal	Equal	Decreased	Equal

Figure 2.1. Internal standard employed in DIGE improves accuracy of quantitation by revealing gel to gel variations. Images represent the perceived abundance of each protein spot in a hypothetical three channel DIGE experiment. Table below the images describe how each protein spot would be interpreted with and without an internal standard.

Analysis of DIGE exeperiments

The traditional method of using one biological replicate per gel, results in a large number of images that must be matched and analysed. One way of minimising variation in these gels is to run replicates at the same time, using the same equipment, however practical considerations make this difficult. By multiplexing samples, the number of gels required is immediately halved, and also ensures that spots can be matched within each gel easily as migration of proteins in each sample/fluorescent channel is identical. Numerous software packages have been developed, by several companies that largely share similar capabilities and functions. We have used DeCyder 6.0 (GE Healthcare) for the analysis of our gel images due to its easy integration with our hardware and cost effective licensing. Gels are scanned at the three wavelengths of the dyes, optimising the exposure time to increase the dynamic range of the signal (the Ettan Imager is capable of pixel values between 0 and 65'500) and minimizing the signal-to-noise (s/n) ratio. Images are subsequently loaded into DeCyder, gel images are normalised between the internal standards, automatic spot detection is performed, spot patterns are matched and modulation of protein is calculated along with statistics for each protein. This workflow is illustrated in Figure 2.2. Referencing the identities of the protein spots back to the gel image can also often provide information about post translational modifications such as phosphorylation, cleavage and alternative splicing.

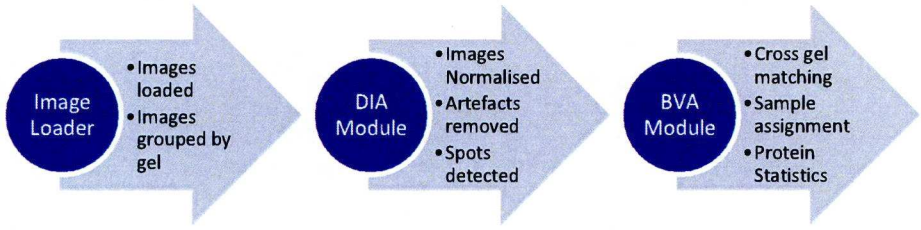


Figure 2.2 Modules, functions and workflow in the analysis of DIGE experiments using DeCyder 6.5.

Image acquisition and analysis method

Images were acquired using an Ettan DIGE Imager (GE Healthcare) and Ettan Imager software, according to the manufacturers standard protocol, which maximises the dynamic range (pixel values ranging from 0 to a possible 65500) of the image without losing quantitative information (via "clipping" of the signal). To quickly determine appropriate exposure times, a small area of the gel expected to contain the most intense protein spots (based on small scale tests) was scanned at each of three wavelengths corresponding to each of the three CyDyes (or Sypro Ruby when scanning a preparative gel). Using the "report view" page of the Ettan imager software the cursor was positioned over the most intense area of the acquired image to obtain a pixel intensity value. The exposure time for each channel of the image was increased or decreased to obtain maximum pixel intensity values of between 30,000 and 50,000. When appropriate values were obtained for each channel, these values were used to scan each gel in its entirety.

To ensure better automatic matching, areas around the edges of the scanned images not containing protein spots or heavily obscured by dye fronts were cropped from the images, imported into DeCyder^{TM} Software (GE Healthcare) for spot detection and matching algorithms. Where the automatic matching was unsuccessful, manual landmarking was performed and used to repeat the spot matching. Levels of image filtration were also applied to remove interfering signals such as dust particles from subsequent analysis. Dust could be reliably detected as "spikes" with a steep slope and small area.

Gels (and hence protein spots) were assigned to their correct treatment group and protein spots with significant changes in their standardized log abundance ($-1.2 \geq$ fold change ≥ 1.2, Student t-test $p \leq 0.05$) were marked for picking from the preparative gel. The preparative gel was post-stained with Sypro Ruby (Invitrogen) in order to identify areas of greatest spot intensity (which differs from areas of greatest CyDye signal by a mass decrease equivalent to the mass of a CyDye molecule).

Spot picking and tryptic cleavage method

Spots assigned for picking were assigned to a 'pick list', a text file containing the x,y coordinates of spots to be excised, relative to a pair of reference markers. This file was uploaded to the Spot Picker (GE Healthcare) robot, which was then used to excise spots from the gel and store them in 96-well plates.

The following spot picking parameters were used in order to minimise loss of gel plugs:

Parameter	**Amount**
Prefill volume	100 µl
Aspirate volume	50 µl
Aspirate flow rate	20 $ml.min^{-1}$
Dispense volume	150 µl
Dispense flow rate	30 $ml.min^{-1}$
Jazz (side-to-side oscillation)	1.1 mm

Plates containing gel 'plugs' were sealed and frozen until they were subjected to tryptic digestion.

After thawing, water dispensed with the plugs during picking was removed with a pipette tip. Plugs were destained for 10 min at 37 °C with 25 mM ammonium bicarbonate, 50 % (v/v) acetonitrile (ACN). All liquid was then replaced by 100 % (v/v) ACN for 10 min at 37 °C. Solvent was then removed and the plug allowed to air dry for 10 min. Ten µl of Trypsin solution (25 µg trypsin diluted with 250 µl 50 mM acetic acid then diluted 1 in 10 with 25 mM ammonium bicarbonate) was added and incubated at 37 °C for 1 hr followed by the addition of a further 10 µl 25 mM ammonium bicarbonate and incubation at 37 °C overnight. Digestion was stopped with 2 µl of 2.6 M formic acid.

Mass Spectrometry using ESI LC MS/MS (LTQ)

This study has used a highly sensitive and accurate mass spectrometer throughout. The system used employs a nano LC separation of peptides which ensures minimal diffusion of reversed phase separated peptides during transport to the electrospray ionisation (ESI) source of the mass spectrometer. In ESI, ionisation is achieved by applying a high voltage to the tip of a very fine capillary tube, charging the surface of emerging droplets. These repelling charges cause disruption of the surface tensions and lead to the evaporation of solvents and breakdown of the droplets leaving charged ions in the gas phase [87]. The invention of ESI is attributed to John B. Fenn, for which he was later awarded the Nobel Prize. The gas phase ions are drawn into the MS where

they are contained by a linear quadrupole ion trap [88] and can be selectively released from the trap according to their mass to charge ratio (m/z). Once a peptide ion has been identified it is further fragmented in the ion trap and the resulting fragment ions, whose masses correspond to the amino acid sequence of the peptide, are measured.

Identification of proteins by tandem mass spectrometry

Twenty µl of the tryptically digested solutions of peptides were transferred to wells in 96-well plates, and increasing volumes of water (Chromanorm for HPLC, VWR, England) were added to compensate for the evaporation that might occur over the course of the analysis based on experimentally determined rates of evaporation at 4 °C (Sanderson, S. J., personal communication). Tandem MS data were acquired for the peptides using a linear ion trap mass spectrometer (MS) equipped with an electrospray ionization (ESI) source coupled online to a Dionex u3000 HPLC utilising a c18 reversed phase chromatography column for separation. Fifteen µl of peptide solution were loaded onto a C18 TRAP column for concentration and desalting, and subsequently loaded onto a c18 reversed phase column equilibrated with 97.9 % H_2O, 2 % acetonitrile (ACN) (v/v) and 0.1 % formic acid (FA) (v/v), at 300 $nl.min^{-1}$. Peptides were eluted using a gradient of ACN, increasing from 2 % ACN (v/v) to a final concentration of 90 % ACN (v/v) (90 % ACN, 9.9 % H_2O, 0.1 % FA (v/v)). Peptides were passed through a PicoTip and ionised at 1.8 kV. Ionised peptides were analysed using the "Big

Three – Triple Play" method which consists of three main stages, run in an iterative manner.

1) Survey scan – A range of m/z values were quickly scanned for the presence of peptides and the top three most intense peaks at a given time point above a threshold of 20'000 were noted for analysis in step 2
2) Zoom scan – For each of the above peaks, a high resolution scan over a narrow m/z window was analysed to determine the precise mass/ charge, the charge state of the ion and the peak collision energy for fragmentation in step 3
3) MS/MS – using the information from step 2, each peptide was fragmented to produce a series of ion masses corresponding to different randomly split peptide fragments. The differences in masses, correspond to the mass of the amino acids in the chain, and the sequences can therefore be calculated from the acquired spectra.
4) Repeat steps 1–3 – Peptides analysed in the current round were excluded from analysis for the next 60 seconds and therefore excluded during the next iteration of analysis

The results of the analysis were stored in .RAW format, and required further processing before they could be used for further analysis. The frequency of ms/ms scans over time was assessed using BioWorks Browser (Thermo Scientific) and areas not containing useful data were excluded from subsequent generation of .DTA files using Sequest [89]. .DTA files were then concatenated using merge.pl, a perl program that converts .DTA files into the

mascot generic format (.MGF) ready for searching against the appropriate database. The resulting MGF file for each sample was searched against the hypothetical ion masses as determined by theoretical tryptic digestion of the *G. lamblia* genome sequence that can be found at G. lambliaDB[90] using the MASCOT search engine [91]. Database searches were performed using a dedicated in-house server, either individually using the standard ms/ms web browser based interface or using the Mascot Daemon to automate the search. In either case the following non-default parameters were used:

Parameter	**Setting**
Database:	G. lamblia Annotated Proteins
Fixed modifications:	Carbamidomethylation (C)
Variable modifications:	Oxidation (M)
Enzyme	Trypsin
Maximum missed cleavages	1
Peptide charge	1+, 2+ and 3+
Peptide tolerance	± 2.0 Da
MS/MS tolerance	± 0.8 Da

An overall Mascot score of at least 50, and at least 70 % presence of b or y ion series in any peptide was required for a positive match.

Aims of this study

This study aimed to use differential proteomic, transcriptomic and bioinformatic techniques to profile the global cellular changes that occur in host and pathogen cells in response to NTZ. It was hoped that these changes could be used to gain novel insights into the mechanism of action of NTZ. Specific molecular interactants with the thiazolides and details of how these interactions may occur are also sought.

In Chapter 3 of this study, transcriptomic, proteomic and bioinformatic approaches are used to investigate changes in a NTZ resistant strain of *Giardia lamblia.* In Chapter 4 the DIGE technique is applied to a human cell line expressing a subgenomic replicon of HCV, which has relevance to the new hypotheses regarding NTZs efficacy being mediated through host cell targets. Benefiting from excellent functional annotation, bioinformatic approaches are also employed to describe the possible implications of the results of this experiment. Global profiling techniques can only go so far to describing the specific interactions of a drug and its targets, and therefore to attempt to answer important questions about specific host cell targets of the thiazolides, affinity chromatography is used in Chapter 5 to seek out proteins with binding interactions with the thiazolides. The potential of this technique for discovering potentially important information is clearly demonstrated by the identification of a previously unidentified binding partner from human liver cells, NAD(P)H Quinone Oxidoreductase 1 (NQO1). In Chapter 6, the nature of this interaction is investigated in greater detail.

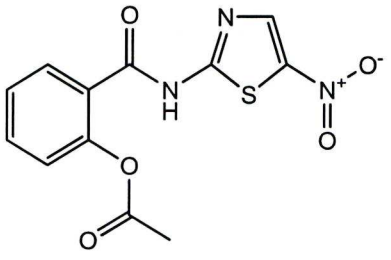
Chapter 3

Proteomic and Transcriptomic Analysis of Resistance to NTZ in *Giardia lamblia*.

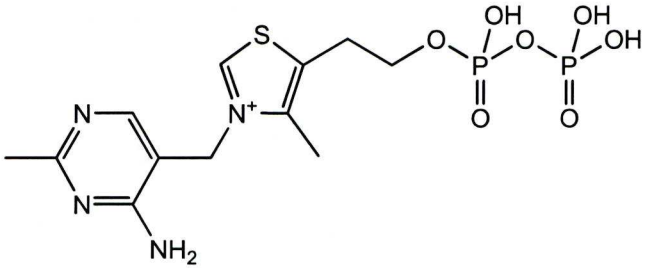
Introduction

The Centre for Disease Control (CDC) in the U.S.A. estimates that there are approximately 8 incidences of *G. lamblia* infection per 1000 people in the U.S.A. every year, and the Health Protection Agency (HPA) reports a similar proportion of reported cases in England, Wales and Northern Ireland. It is not in developed countries however that giardiasis represents a major clinical problem. In third-world countries, the prevalence of infection can be over 50 % in some populations, putting the world-wide total in the order of hundreds of millions (approx. 250 million infections, 500000 symptomatic according to World Health Organization estimates for 1992) [92]. In these areas, death resulting from dehydration and malnutrition due to diarrhoea is the main cause for concern. There have also been suggestions that chronic giardiasis can result in long-term growth retardation [93]. Thankfully, there are several pharmaceuticals that are commonly used to treat giardiasis, including metronidazole, tinidazole, nitatazoxanide, and less commonly paramomycin, furazolidone, quinacrine, albendazole and mebendazole. Currently however, treatment of giardiasis is largely by metronidazole (also called Flagyl™). Metronidazole is believed to act against anaerobic bacteria and susceptible protozoa by reduction of its nitro group by pyruvate ferredoxin oxidoreductase (PFOR) (the homologue of aerobic pyruvate dehydrogenase enzymes), the

terminal part of the glycolytic pathway. This generates reactive toxic nitro radicals that cause damage to the cell. The drug is generally well tolerated, although some patients suffer with unpleasant side effects and incidences of resistance are occasionally reported. In laboratory derived resistant isolates of *G. lamblia*, PFOR is often downregulated several fold and complementary enzymes are believed to take its place. Nitazoxanide also possesses a nitro group on its thiazole ring, raising the possibility that a mechanism of action similar to that of metronidazole could be responsible for its anti-*G. lamblia* activity [94]. Research however showed that no such metabolism of NTZ could be observed [24]. Parallels were also drawn between the cofactor of PFOR, thiamine pyrophosphate (TPP – Figure 3.1), which shares some structural similarity to the thiazolides. It was postulated that NTZ might compete with TPP for a binding site on PFOR, thus inactivating or modifying its activity. However, NTZ appears to hijack a TPP intermediate and cause the reversal of pyruvate binding [24]. This mechanism is thought to have a high barrier to resistance since mutations in PFOR would not affect the interaction of TPP and NTZ without affecting the interaction between PFOR and TPP, and could have lethal effects.



Nitazoxanide



Thiamine

Figure 3.1. Structures of nitazoxanide and thiamine pyrophosphate.

Application of proteomics to the discovery of mechanisms of action

The discovery of the mechanism of action of a drug can often be expedited by the judicious use of proteomic or transcriptomic techniques and protein expression profiling can reveal important changes in the abundance of proteins or metabolic pathways involved in conferring resistance to a drug. Studies of the *Leishmania* proteome using 2D gel electrophoresis (2D-GE) have revealed over-expression of known drug resistance factors and have also postulated post-translational modification of drug targets that had not been previously considered [95]. The use of 2D-GE has since been improved greatly and techniques such as difference gel electrophoresis (DIGE) have gained popularity in investigating resistance and virulence in many different organisms. Recent examples include chemoresistance in cancer cell lines [96], drug resistance in *Candida albicans* [97] and *Acinetobacter baumannii* [98] and virulence factors in *Entamoeba histolytica* [99] and examples for many other pathogens exist.

The speed with which custom high resolution oligonucleotide arrays can now be generated for organisms with sequenced genomes, has meant that comparative array techniques have now been used to discover mutations that are associated with drug resistance [100]. High throughput sequencing methods are also now beginning to provide an accurate and rapid utility for profiling changes in transcript expression and for discovering mutations [101, 102].

Suitibility of G. lamblia for proteomic and transcriptomic study

Aside from its interesting phylogenetic positioning that indicates that it is a very early branching eukaryote, one key factor that makes *G. lamblia* an ideal candidate for study is the ability of some isolates to be grown either axenically, or in co-culture with a host cell. This provides scope for experimental approaches that ensure that observed protein changes can be directly attributed to the perturbation of the system under consideration, and are not indirect effects mediated through the host system for example, as might be the case with other parasites such as *Toxoplasma gondii*. The requirements of a media for axenic growth mimic the favourable conditions of the host [103] including low O_2 content, high cysteine content and the addition of bile salts. Modified TYI-S-33 is the most commonly used growth media in the culture of *G.lamblia*. There are however several other factors and technological developments that make *G.lamblia* increasingly appealing for a model for study.

G. lamblia Genomics

A major technological advance that has enabled more effective research into the *G. lamblia* proteome is the availability of a fully sequenced genome (available at www.giardiadb.org [104]). Whilst at the start of this project, the G. lamblia genome database was at a fairly early stage of development, considerable efforts of researchers in this area has meant that the genome is now fully sequenced

with good coverage and furthermore it is provided alongside a range of tools for querying the data. The data has also been integrated with another cross species database of related organisms, EuPathDB (www.eupathdb.org), which provides several tools for cross species comparisons. Perhaps most importantly, the availability of this database facilitates the use of other techniques that rely on these data, i.e. transcriptomics and proteomics.

G. lamblia Transcriptomics

The availability of comprehensive genomic sequence data makes the use of microarray technologies for the study of *G. lamblia* now feasible. The Pathogen Functional Genomics Resource Centre (PFGRC) (http://pfgrc.jcvi.org/) has free arrays available but approximately 50 % of the oligomers on these arrays belong to now depreciated gene sequences. Thus far, the use of these and other arrays appears to be largely limited to attempts to correctly diagnose *G. lamblia* as the causative agent in gastrointestinal disease where it is present. They have also bee used to investigate unresolved questions about *G. lamblia's* phylogenetic classification [105]. One notable exception was as study performed by researchers at the University of Berne, Switzerland [106]. In their study, microarrays were used to investigate laboratory derived resistance to NTZ. Another transcriptomic study worthy of note has looked at host cell changes in response to infection with *G. lamblia* [107]. The results of this study suggest that *G. lamblia* causes many changes in expression in host cells and in particular *G. lamblia* appears to induce chemokine responses different to other enteric pathogens that cause intestinal inflammation, providing a possible insight as to

how *G. lamblia* establishes chronic infection, often without causing significant inflammation in humans.

G. lamblia Proteomics

In *G. lamblia* the level of refinement and annotation of the genomic data is improving rapidly and a recent release of the database (version 1.2) eliminated nearly half of the predicted genes on the basis that they were unlikely to represent true proteins. However, current gene annotation is heavily reliant on homologous proteins from other organisms and the multiple and sometimes unexpected functionalities of some proteins, means that species specific functional studies are needed for accurate interpretation of results. Improvements in *G. lamblia* specific data would also create a better foundation for the creation of *in silico* models. Currently only a relatively small proportion of genes have functional annotations associated with them, and *G. lamblia*-specific proteins such as the giardins might never be described fully without direct experimentation.

There are relatively few existing studies that have investigated protein expression in *G. lamblia*, and certainly none that investigate how protein expression might be modified in response to NTZ resistance. The majority of studies are focused towards laying more complete foundations of our understanding of the biology of the parasite or understanding the way the parasite interacts with its hosts. A recent review [108] gives an excellent overview of this area, and also highlights ways in which the development of this area could be improved. It praises the integration of the mature genomes of

parasites such as *Toxoplasma gondii* with thorough proteomic profiling studies that have used multiple techniques to identify up to 30 % of the predicted proteome [109].

Aims of study

We wish to improve the understanding of the mode of action of nitazoxanide against *G.lamblia* parasites. *G. lamblia* can be grown axenically and its genome has been fully sequenced and partially annotated. Furthermore, a laboratory derived, NTZ resistant strain (also cross resistant to MET) and a MET resistant strain are available for comparison with the wild type one (NTZr/METr, NTZs/METr, and WT respectively). The study will primarily focus on the problem from a proteomic perspective, but will also be followed up using both quantitative transcriptomic data, and also with the re-analysis of an existing microarray dataset. Together these approaches will be used to investigate how gene expression is modulated in parasites that have acquired resistance to NTZ and MET, and the results used to infer possible mechanisms of resistances.

Methods and Materials

Description of preparation of resistant strains

Samples of *G. lamblia* were kindly provided by Prof. Andrew Hemphill at the Institute für Parasitologie at the University of Bern and were grown as follows: Trophozoites from *G. lamblia* WB clone C6 were grown under anaerobic conditions in 10 mL culture tubes (Nunc, Roskilde, Denmark) containing modified TYI-S-33 medium [110]. To initiate subcultures, cultures with confluent trophozoite lawns were incubated on ice for 15 min. Suspended living (adherent) trophozoites were counted (Neubauer chamber, 200x). Subcultures were initiated by adding 10^4 trophozoites to a new culture tube. The METr clone was cultivated in the presence of concentrations of metronidazole increasing to 60 μM, whereas double-resistant clone NTZr/METr was cultivated in the presence of up to 40 μM nitazoxanide [106]. Cells were passaged in the absence of drug for several generations before being snap frozen in nitrogen ready for transport.

2-Dimensional difference gel electrophoresis (DIGE)

A 2D-DIGE experiment was performed using the common methods outlined in Chapter 2 and the experiment specific manner below.

Cell Lysis and Sample Preparation for Proteomic Analyses

For each of the cell lines four biological replicates were obtained and each frozen pellet was solubilised in cell lysis buffer (20 mM Tris, 7 M urea, 2M

thiourea, 2 % (w/v) CHAPS, 2 % (w/v) ASB-14, complete mini protease inhibitor, DNase, pH 8.5 (on ice)) using 1 ml per 1.1 x 10^8 cells.

Total protein concentration was determined using the 2D-Quant kit (GE Healthcare) in which proteins are precipitated to leave potential interfering contaminants in solution followed by resuspension in a copper containing solution and a quantitative measurement of the level of unbound copper that remains after binding to protein has occurred. The 2D Quant Kit is compatible with all of the reagents in the lysis buffer except ASB-14 which had not been tested. The results of the assay proved to be unreliable possibly as a result of the ASB-14. To overcome this, samples were precipitated to remove salts, contaminants and other interfering substances using the 2D Clean-Up kit (GE Healthcare) using the most recent concentration values obtained for the assay and resuspended in 8 M urea, 30 mM Tris, 4 % (w/v) CHAPS, pH 8.5 so as to be compatible with the BSA-coomassie assay (Bio-Rad). After determination of protein concentration, all samples were adjusted to the same concentration (equal to the lowest measured concentration obtained) and stored frozen.

CyDye™ Minimal Labelling

The following nomenclature was used for each sample, WT, C4 and C5 refer to the Wildtype, NTZr/METr and NTZs/METr strains respectively and subsequent numbering refers to the biological replicate. Labelling was performed as outlined in table 3.1 and 3.2 using the method described in Chapter 2.

Gel Id	Cy3	Cy5	Cy2
1	W3	C41	STD
2	W2	C44	STD
3	C42	C52	STD
4	C51	W1	STD
5	C53	C54	STD
6	C43	W4	STD

Table 3.1. Labelling and "Gel Id" allocation of samples. W, C4 and C5 refer to the WT, NTZr/METr and NTZs/METr strains respectively and subsequent numbering refers to the biological replicate. The internal standard (STD) was created from equal amounts of all samples. An unlabelled preparative load was added to Gel 3.

Gel Id	Cy3	Cy5	Cy2	Sypro
1	W3	C43	STD	-
2	W2	C44	STD	-
3	C42	W4	STD	-
4	C41	W1	STD	-
5	-	-	-	Prep

Table 3.2. Labelling and "Gel Id" allocation of samples. W and C4 refer to the WT and NTZr/METr strains respectively and subsequent numbering refers to the biological replicate. The internal standard (STD) was created from equal amounts of all samples. An unlabelled preparative load run on a separate gel (Gel 5).

qRT-PCR

To provide accurate and useful information about the underlying cellular processes controlling protein expression we measured levels of mRNA transcripts of 22 genes encoding the proteins differentially expressed (as determined by 2D-DIGE) between wildtype and NTZr/METr strains of *G. lamblia.*

Sample preparation

cDNA and RNA samples were prepared at the Institute für Parasitologie by Carolin Maier, University of Bern using the Qiagen RNeasy^{TM} kit and DNase 1 for the RNA isolation and the Quiagen Omniscript^{TM} RT kit for reverse transcription to cDNA, according to manufacturers' instructions in both cases. A polyT-ANC primer was used for subsequent real-time PCR.

qRT-PCR procedure

Quantitative real-time PCR was performed with 4 µL of 1:100 diluted cDNA using the Quanti TectTM SYBR Green PCR Kit (Qiagen) in a 10 µL standard reaction containing a 0.5 mM concentration of forward and reverse primers (complete list in Table 3.3)(MWG Biotech, Ebersberg, Germany). Furthermore, a control PCR included RNA equivalents from samples that had not been reverse transcribed into cDNA (data not shown) to confirm that no DNA was amplified from any residual genomic DNA that might have resisted DNase I digestion. PCR was started by initiating the 'Hot-Start' Taq DNA-polymerase reaction at 95°C (15 min). Subsequent DNA amplification was performed in 40

cycles including denaturation (94°C for 15 s), annealing (60°C for 30 s) and extension (72°C for 30s); temperature transition rates in all cycle steps were 20°C/s. Fluorescence was measured at 82°C during the temperature shift after each annealing phase. For each gene, three replicate experiments were performed. Expression levels of the genes summarized in were given as values in arbitrary units relative to the amount of constitutively expressed housekeeping gene ACT [106].

Gene Target	ORF number	Forward/ Reverse Primer	Primer Sequence
Hypothetical protein	4259	F	TGACATGTTCGGTTTTGGGG
		R	AGCAGCAGGCTGCTGGGT
Beta-giardin	4812	F	AGGTTCCACGACAAGATGGA
		R	GATGGCGATCTGGTTCTGG
Calmodulin	5333	F	CCTATAAACCTGCCCCAGG
		R	CATTGACGTACCCGTCACC
Hypothetical protein	8038	F	TTTGGACTCGGCAATACCCA
		R	TCTTCGAAAGGCTGGTTTCC
Protein disulfide isomerise 5	8064	F	TGCATCTCAACACGTCATATG
		R	CCTCCTGGCAGACCTCAC
Nucleoside diphosphate kinase	11301	F	TGAACGCACCTTTCTCATGG
		R	CCCCATCATGGTGCGGGA
20S proteasome alpha subunit 2	11434	F	AGGAAGGCTCTCAATAGGGA
		R	ACGAGCTAACATATTAACGGG
Hypothetical protein	11497	F	CCCCCTGAATTTCAACCAGA
		R	TGAGAGACGTGATGGTGGG
Carbamate kinase	16453	F	CTTAGCGAGATTCTCGCCC
		R	ACGAAGGTAGAGCAGCTGG
Hypothetical protein	16588	F	AGCACCTTGCGGCGTACC
		R	TCTTGGCTGCTGCAGGGG
Protein 21.1	17060	F	CCCCTGCAGCAGCCAAGA
		R	CCGTTAGGGCTGCTTGGG
Giardia trophozoite antigen 1	17090	F	ACTCCTGTTGCAAAGATATGG
		R	TATCTCCCCGGAGTTCTCC
PPI-B precursor	17163	F	AAGGTCTTCTTTGACATCACC
		R	GCCCAGCATTCGCCATGG
Gamma giardin	17230	F	GTCATCGTTCTCGACCGTC
		R	GTAGACTCCAGCTCGCGG
Xaa-Pro dipeptidase	17327	F	CCTCTACGTCACGGGCTG
		R	CCCCGCCGTCTGCTCCA
Elongation factor 2	17570	F	GGTTGTGATGGAGGAAGCC
		R	GTCTGGACGCAAACACCCT
Spindle pole protein, putative	24537	F	CTTAAGGTTGTGGCTCCCC
		R	TGCCTGCAGTGAAGCGCC
Protein 21.1	27925	F	GAGTGTGTGAAGATCCTGGA
		R	AGGCAATCAGTTGAAAAGGGA
Delta giardin	86676	F	GCGCGCCTTCTGCCCTC
		R	TTGTCAAGGTGCTCAATCCC
Triosephosphate isomerase	93938	F	TGCCTGCTCGTCGCCCC
		R	CCCCGTTCCCCTCTAGGT
Alpha-7.3 giardin	114787	F	GCTGCAGCGAAGGCTACC
		R	GATCAGCCTGCAGCGGAC
Protein 21.1	14872	F	AACAGGCACCTACAACCCC
		R	TTAGAATAGTTTCCCCCACCA

Table 3.3. Primers used for qRT-PCR experiments. Open reading frame (ORF) numbers correspond to ID's for species GL50803(WB C6) in GiardiaDB (www.giardiadb.org)

Microarray Method

Transcriptomic analysis of resistance to NTZ in Giardia lamblia

To gain further insight into gene expression changes that are associated with resistance to NTZ, a microarray dataset originally created at the University of Bern was re-analysed. This two colour microarray dataset using the NTZr/METr cell type from the same source has thus far only had relatively crude and arbitrary criteria of >-4, <4 fold change applied to it to determine potentially altered gene expression in the drug resistant cell type. We have applied more accurate and statistically robust methods to the analysis of these data including normalisation that accounts for inherent systematic spatial effects (i.e. print tip group effects) as well as investigating the use of other more broad methods such as over-representation analysis, using hypergeometric distibution of Gene Ontology terms.

The experimental work for the microarray experiments is entirely the work of researchers at The University of Berne, Switzerland. Detailed descriptions regarding the preparation of RNA samples, microarrays, hybridisation and data acquisition can be found in the original paper [106].

A brief description of the microarray method is reproduced here, with permission, for convenience.

Aminoallyl cDNA was prepared using a combination of random primers (Promega) and Omniscript reverse transcriptase and was subsequently purified using a Roche PCR purification kit to remove contaminating amino groups that would interefere with subsequent labelling. Samples were then lyophilised

until required. Lyophilised samples were rehydrated and labeled with either Cy3 or Cy5 cyanine flourescent dyes. Dye-labelled cDNA was then purified by Roche PCR purification kit according to the protocol. Purified cDNA was again lyophilized and stored in the dark until microarray hybridization. Microarrays from the Pathogen Functional Genomics Research Center (J. Craig Venter Institute, Rockville, MD, USA) containing 19230 elements consisting of duplicates of 70-mer oligomers derived from 9115 predicted open-reading frames (ORFs) including the clearly identified 6470 ORFs of the genome of *G. lamblia* WB C6 as described in the *G. lamblia* database and of 500 *Arabidopsis thaliana* control oligomers were prepared for hybridisation. Slides were hybridised with samples, pooled from equal amounts of the labelled cDNA, washed and subsequently scanned using a GenePix Personal 4000B scanner (Molecular Devices, Sunnyvale, CA, USA). Data files including annotation and evaluation of the hybridized slides was performed using the GenePix Pro. Two technical replicates were performed in total.

Re-analysis of WT vs. NTZr/METr microarray data

As part of this study, to gain further insights into the transcriptomic data from the microarray data, it was re-analysed using the statistical computing tool R [111] (www.r-project.org). The analysis also used the following package libraries which provided the means to execute specific functions of the analysis: limma, limmaGUI, statmod, R2HTML, tcltk, marray and sma [112-115]. The data used in this analysis is available online at the following location:

http://www.ipa.vetsuisse.unibe.ch/index.php?option=com_content&task=view&id=56&Itemid=134&lang=EN.

The data available is a Excel spreadsheet (Microsoft Corporation) copy of the output from the microarray analysis software Genepix Pro (Molecular Devices), and essentially contained the same information as a .gpr file (the format used by Genepix) without the header information that described some of the scanner settings etc. In order to make the data contained in the .xls file compatible with the .gpr format, the raw tab delimited data was edited in Notepad++ and the header information was added to the file. Once raw median foreground and background expression levels for each colour channel (Cy3 (532 nm) and Cy5 (635 nm)) were loaded into limma by LimmaGUI, the data was background subtracted using the normexp + offset method (offset =50) rather than local median feature background subtraction which often results in a high number of negative expression values which later results in missing data due to the inability to calculate the $\log_2$ ratio from a negative value.

Within each array, the data was normalised using a print-tip group loess normalisation algorithm and the data did not require normalisation between arrays. Within array duplicates were averaged before applying a linear fit model and genes were ranked in order of the significance of their evidence for differential expression using the eBayes method (empirical Bayes) to compute moderated t-statistics, moderated F-statistic, and log-odds of differential expression and sorted by their log odds with the topTable function. P-values were adjusted using the "Holm" method, a less stringent step-down modification of the Bonferroni method, however since current available methods cannot take

into account the likely positive correlation both within array and between array technical replicates and the sample size (i.e. two technical replicate arrays) is very small, this was merely used for indicative and illustrative purposes.

Data were exported back to Excel and spots with low expression levels (<800) or $\log_2$ fold-change less than ±1 (actual ±2 fold change) were excluded. Product descriptions were also updated using the latest release of G. lamblia DB (which saw a large proportion of the 9115 original set of G. lamblia genes included in the microarray experiment depreciated leaving only 4692 genes). Available gene ontology terms (GO terms) were also extracted from the database and used for overrepresentation analysis.

Bioinformatic Analyses

Informatic analysis of DIGE experiments

To try and deconvolute protein expression changes due to NTZ resistance, from generic stress responses, or changes that confer resistance to MET, the results from the WT vs. NTZr/METr analysis were combined with the WT vs. NTZr/METr+NTZs/METr combined experiment. Both of the datasets have the potential to answer different biological questions, i.e. the comparison between the WT and NTZr/METr strain, should detect modulated proteins that result in resistance to either NTZ or MET or both but may also contain information that relates to a stress response. The WT vs NTZr/METr+NTZs/METr comparison however should mainly detect modulated proteins that result in resistance to MET since both a NTZs and NTZr strain are included in the analysis. A stress response may also be detected. Thus by using the latter data set to infer which proteins from the first data set might be modulated due to METr or a stress response it is hoped the power to identify proteins specific to NTZ resistance will be improved. Furthermore, by describing the data in a more generic way using Gene Ontology (GO) terms, it was hoped that broader areas for investigation might be identified.

Proteomic Analyses

GO Terms

Proteins found to be differentially expressed by DIGE were interrogated for function using GO term annotations. Gene and GO term data were downloaded from G. lambliaDB and from this an annotation file compatible with

Ontologizer was created. Ontologizer's integrated GraphViz connection was used to generate a graph of the full lineage of all GO terms found in the proteomic datasets using the default graph rendering tool 'Dot'.

Transcriptomic Analyses

GO-Term Hypergeometric Distribution Testing

Gene and GO term data were downloaded from G. lambliaDB and from this an annotation file compatible with Ontologizer 2.0 was created. When provided with a population and its respective annotations, Ontologizer was used to determine overrepresentation using hypergeometric distribution analysis, initially using the parent-child setting in order to account for the hierarchical nature of GO terms but also using term-for-term and topology-weighted methods. Ontologizer's integrated GraphViz connection was used to generate a graph of the significant GO terms using the default graph rendering tool 'Dot'.

Results

A direct proteomic comparison between the WB C6 wildtype strain of *G. lamblia* (WT) and a laboratory derived NTZ resistant (NTZr) cell line [28] was initially conceived. A challenge in interpreting this approach is the inherent difficulty in determining which changes in protein abundance are the direct effects of the characteristic under study (i.e. drug resistance in this case). For example, a recent proteomic investigation of encystation in *G. lamblia* [116], indicated the involvement of HSPs 70 and 90, however due to the acidic bile conditions required to induce encystation, it is difficult to determine if these proteins are essential to encystation, or if they simply facilitate survival. To resolve this dilemma, in this study two protein data sets were created. The first compared the WT to the NTZr/METr strain, and a second dataset was generated that included both the NTZr/METr strain and a metronidazole resistant (NTZs/METr) strain. It was hoped that by comparing these datasets we might be able to deconvolute, the effects of drug stress, and also detect proteins whose modulation might be more relevant to MET resistance than NTZ resistance in the cross resistant strain.

Proteomic analyses

2D-DIGE of Wildtype vs NTZr/METr Giardia lamblia.

To investigate potential protein changes that might have occurred during the acquisition of resistance to NTZ, cell lysate from wildtype and NTZr/METr *G. lamblia* were compared using 2D-DIGE.

Typical 2D gels

Small scale tests of the lysis conditions using a mini SDS PAGE gel and both pH 3-10 NL and 4-7 IEF strips were successful (not shown), so 3-10 NL strips were used for large scale gels to maximise proteome coverage. A typical separation is shown in Figure 3.2, including an overlay of the three CyDye wavelengths illustrating how protein levels can been seen to differ between samples.

Spot matching and DeCyder analysis examples

Whilst spot detection in DeCyder was comprehensive and generally satisfactory, spot matching between gels usually required significant manual intervention (landmarking) to achieve an acceptable level of accuracy.

Protein identifications

The majority of protein identities in the following lists are assigned on the basis of being the 'top hit' presented by the results of the Mascot scoring algorithm (that is, possessing the strongest statistical evidence for a correct identification), however in two cases, ambiguity in the assignments resulted

from Mascot's scoring system automatically switching its proprietary scoring system to an alternative one upon submission of a query where;

"the ratio of the number of queries to the number of entries in the database, (after any taxonomy filter), exceeds 0.001"[91]

This scoring algorithm is intended for use with MudPIT data however, and is intended to account for protein hits that have arisen as a result of a high number of low scoring peptides. A low intensity threshold during LC/MS to maximise sensitivity and the small size of the *G. lamblia* genome means in many cases the scoring algorithm was changed automatically. To ascertain the extent to which the scoring algorithm affected the assignment of a top hit, the results were re-scored using the standard scoring algorithm with an ion score cut-off of 20. Ambiguity in the assignment of 'top hit' was only a factor in the two aforementioned cases and the result pertaining to Elongation factor 1 was discounted in favour of the default result due to the apparent mismatch between the proteins predicted mass and the observed migration distance. The results and analysis take this into account.

Thirty one differentially expressed protein spots could be matched to a corresponding spot on the preparative gel. Of the thirty-one spots, identifications were made for twenty-five. In each case, the highest scoring protein as determined by Mascot, matching the specified criteria was used for the identification. In total, eighteen unique proteins were detected as being differentially expressed due to the same protein being identified in multiple spots in four instances. Whilst both alpha-7.3 and beta giardin hits were

modulated in the same direction, spots corresponding to protein 21.1 (GL50803_17060) and peptidyl-prolyl cis-trans isomerase B were modulated in both directions. These results might be evidence of an isoform shift or post translational modifications (PTM). In All twenty-seven identifications are shown in Table 3.4.

Simple validation of protein identification results

As another measure of the accuracy of the Mascot hits the predicted protein masses from G. lambliaDB were plotted against the observed migration distance on the gel. The majority of protein identifications aligned well along a single straight line. Proteins that did not align well were identified as Protein 21.1(GL50803_17060 / GL50803_27925) and Elongation factor 2 (GL50803_17570) as indicated by black arrows in Figure 3.3.

Protein Name	ORF Number	t-test	Average Ratio (WT: NTZr/METr)
Beta-giardin	4812	0.015	-1.32
Beta-giardin	4812	0.034	-1.25
Calmodulin	5333	0.002	2.09
Hypothetical (putative nucleoporin)	8038	0.038	1.63
Protein disulfide isomerase PDI5 / Dyenin light chain	8064/9848	0.043	-1.36
20S proteasome alpha subunit 2	11434	0.00069	1.57
20S proteasome alpha subunit 3	14497	0.0043	-1.83
Hypothetical (Bacterial type SMC domain)	15499	0.036	1.84
Carbamate Kinase	16453	0.045	-1.48
hypothetical (ribosomal protein p2)	16588	0.034	-1.32
Protein 21.1	17060	0.0013	2.62
Protein 21.1	17060	0.0036	2.77
Protein 21.1	17060	0.031	-2.97
Giardia trophozoite antigen GTA-1	17090	0.047	1.35
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.012	-1.44
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.014	1.64
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.032	2.48
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.048	1.66
Xaa-Pro dipeptidase	17327	0.035	1.88
Elongation Factor 2	17570	0.026	-1.57
Spindle pole protein	24537	0.033	2.35
Protein 21.1 (Ank repeats / Bacterial type SMC domain)	27925	0.047	-1.37
Triosephosphate isomerase, cytosolic	93938/112304	0.0079	1.57
Alpha-7.3 giardin	114787	0.038	1.51
Alpha-7.3 giardin	114787	0.044	1.77

Table 3.4. Identities and average ratio change for proteins differentially expressed in the NTZr/METr strain of *G. lamblia*. Cell lysates of WT and NTZr/METr *G. Lamblia* were compared using DIGE. Protein spots with significantly different (t-test = <0.05) relative abundance were picked from the gel and identified using mass spectrometry. Duplicate identities in the table arise from the same protein being identified in independent spots.

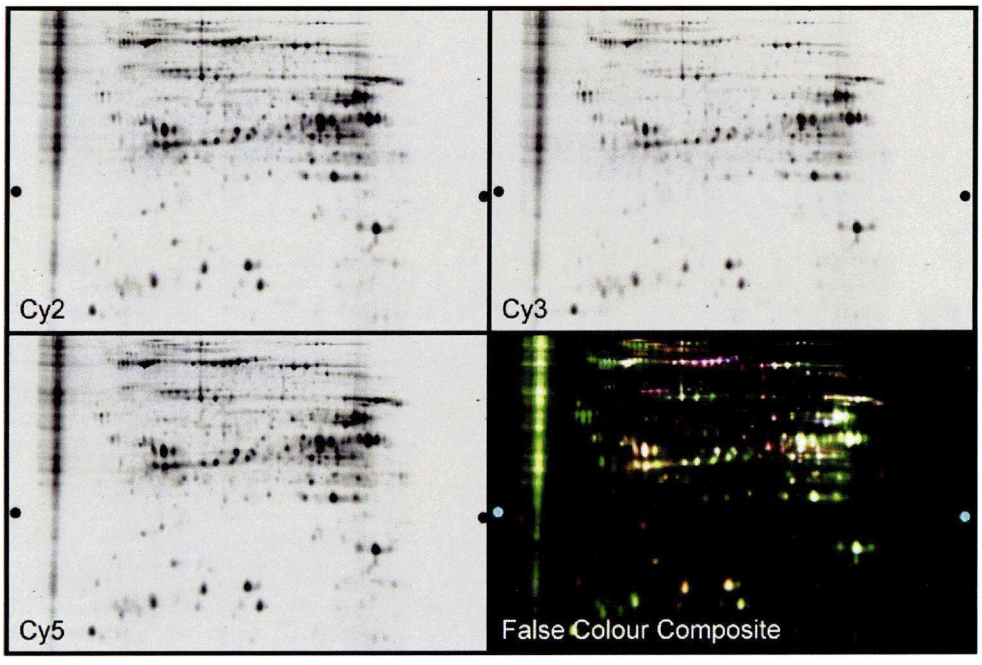


Figure 3.2. A typical 2D separation of *Giardia lamblia* proteins. Panel a) Cy2, b) Cy3, c) Cy5, d) False colour composite of all channels

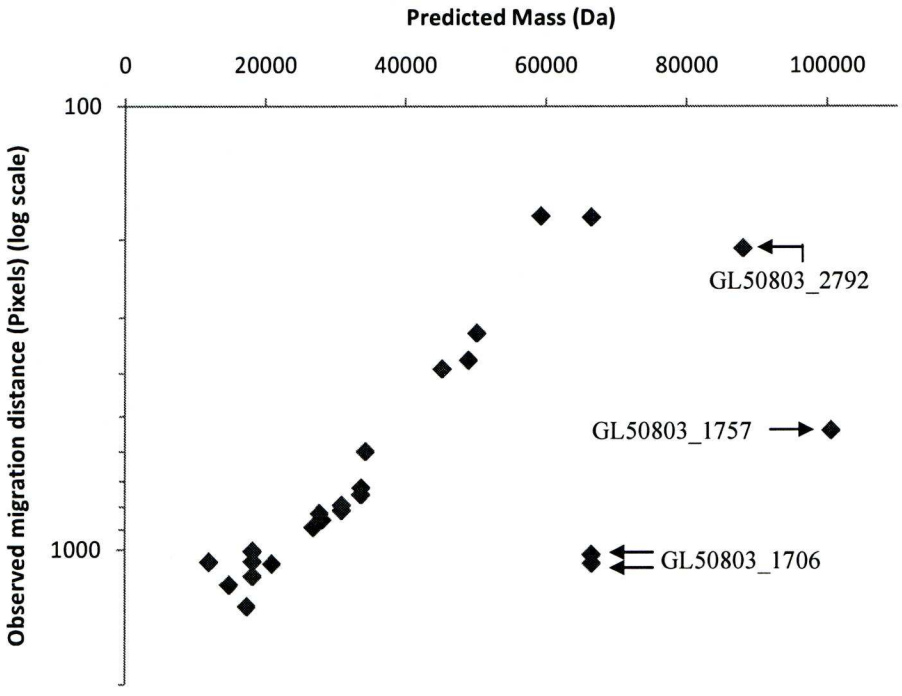


Figure 3.3. Plot of the predicted mass of each protein from release 1.2 of GiardiaDB against observed migration distance on a logarithmic scale. Arrows indicate points corresponding to protein spots identified as Protein 21.1 (i.d.:GL50803_17060 / 27925) and elongation factor 2 (GL50803_17570) that do not appear to correspond to the rest of the data, possibly as a result of either incorrect assignment of identity, alternative splicing or protein degradation.

2D-DIGE of WT vs. NTZr/METr + NTZs/METr Giardia lamblia

Protein identifications

As previously mentioned, the NTZr/METr strain of *G. lamblia* is also cross resistant to metronidazole, however a NTZs/METr strain of *G. lamblia* has also been generated by passaging growing cells in sub-lethel concentrations of MET. In order to attempt to understand the similarities and differences between the resistance mechanisms that occur in each of the strains, a second DIGE experiment was designed to compare the NTZr/METr + NTZs/METr strains together and the WT. Using this dataset, it should be possible to identify the proteins that have changed as a result of being grown under drug pressure, albeit at sub-lethal levels. Moreover, if the mechanisms that provided resistance to the METr strain are similar, or the same as those conveying resistance in the NTZr stain, we might expect to see very similar proteins and processes modulated as in the WT vs. NTZr/METr dataset.

After matching the gel images there were over 200 protein spots that were significantly modulated at the 5 % level, so in order to make the analysis more practical, the protein spots with significant differences (t-test ≤ 0.01) were picked for analysis. The results are shown in Table 3.5.

Protein Name	DB Identifier	t-test	Average Ratio (WT: NTZr/METr+NTZs/METr)
Hypothetical protein	4259	0.004	5.56
Hypothetical protein (general stress/FMN-binding split barrel superfamily)	5810	0.0093	3.11
Wos2 protein	10429	0.003	4.41
Wos2 protein	10429	0.0035	-1.78
Hypothetical protein (bacterial type SMC domain)	13584	0.0049	1.8
Peroxiredoxin 1	14521	0.0021	-1.93
Peroxiredoxin 1	14521	0.0073	2.6
Alpha-6 giardin	14551	0.0064	1.67
Protein 21.1	14859	0.0041	-1.43
Protein 21.1	14859	0.0056	1.52
Peroxiredoxin 1	15383	0.0061	2.33
Protein 21.1	17060	0.0028	-1.38
Protein 21.1	17060	0.0071	2.28
Alpha-11 giardin	17153	0.0086	-1.27
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.006	-1.26
Peptidyl-prolyl cis-trans isomerase B precursor	17163	0.0076	-1.42
Gamma giardin	17230	0.002	1.75
Gamma giardin	17230	0.0078	2.05
Elongation factor 2	17570	0.0017	-1.84
Protein 21.1	27925	0.0037	2.52
NADH oxidase lateral transfer candidate	33769	0.0033	-1.75
Alcohol dehydrogenase	93358	0.0095	2.13
SKD1 protein	101906	0.00015	6.28
Arginine deiminase	112103	0.0015	-1.58
Arginine deiminase	112103	0.0036	-1.74
VSP with INR (TSA417)	113797	0.0045	-5.22
Alpha-7.3 giardin	114787	0.0028	1.82
Alpha-7.3 giardin	114787	0.0051	1.73
Alpha-7.3 giardin	114787	0.0079	1.53

Table 3.5. Identities and average ratio change for proteins differentially expressed in both the NTZr/METr *and* NTZs/METr strain of *G. lamblia.* Cell lysates of WT and NTZr/METr + NTZs/METr *G. lamblia* were compared using DIGE. Protein spots with significantly different (t-test = <0.05) relative abundance were picked from the gel and identified using mass spectrometry. Duplicate identities in the table arise from the same protein being identified in independent spots.

Combined analysis of DIGE datasets

Proteins differentially expressed in both DIGE experiments

To identify proteins more likely to be generic responses to the sub-lethal doses of either NTZ or MET in the growth media the proteins that are shared between both analyses were identified and are given in Table 3.6.

Proteins differentially expressed in only WT vs. NTZr/METr DIGE

Protein lists for both DIGE experiments were also compared to determine which proteins were modulated uniquely in the WT vs. NTZr/METr DIGE in order to enrich for proteins that are related to resistance to NTZ. The results of this are given in Table 3.7.

Protein Name	ORF Number
Protein 21.1	17060
Peptidyl-prolyl cis-trans isomerase	17163
Elongation factor 2	17570
Protein 21.1 (Ank repeats/ Bacterial type SMC domain	27925
Alpha 7.3 giardin	114787

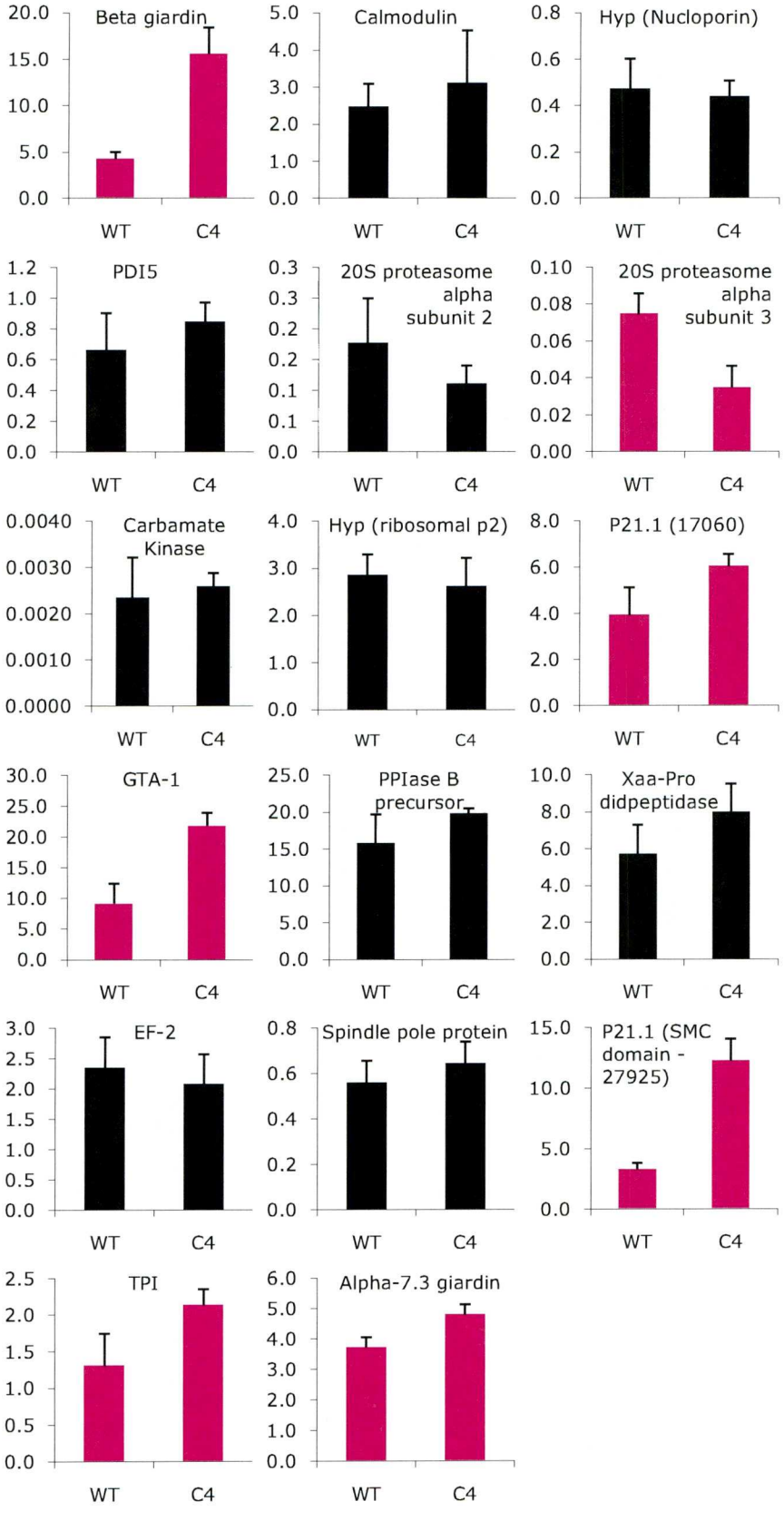
Table 3.6. *G. lamblia* proteins differentially expressed in both WT *vs.* NTZr/METr and WT *vs* NTZr/METr + NTZs/METr DIGE experiments.

Protein Name	ORF Number
Beta-giardin	4812
Calmodulin	5333
Hypothetical (putative nucleoporin)	8038
Protein disulfide isomerase PDI5 / Dyenin light chain	8064/9848
20S proteasome alpha subunit 2	11434
20S proteasome alpha subunit 3	14497
Hypothetical (Bacterial type SMC domain)	15499
Carbamate Kinase	16453
hypothetical (ribosomal protein p2)	16588
Giardia trophozoite antigen GTA-1	17090
Xaa-Pro dipeptidase	17327
Spindle pole protein	24537
Triosephosphate isomerase, cytosolic	93938/112304

Table 3.7. *G. lamblia* proteins differentially expressed in only the WT *vs* NTZr/METr DIGE experiment.

qRT-PCR analyses

To further investigate the detected protein changes identified by 2D-DIGE, levels of mRNA transcripts of the corresponding genes were measured by qRT-PCR relative to the actin (ACT) gene which has been shown to remain unchanged between the wildtype and NTZr/METr clone [28]. The results of these measurements are shown individually for clarity in Figure 3.4 and together to more clearly illustrate relative levels of transcript abundance in Figure 3.5.


Beta giardin
20.0
15.0
10.0
5.0
0.0
WT
C4
Calmodulin
5.0
4.0
3.0
2.0
1.0
0.0
WT
C4
Hyp (Nucloporin)
0.8
0.6
0.4
0.2
0.0
WT
C4
PDI5
1.2
1.0
0.8
0.6
0.4
0.2
0.0
WT
C4
20S proteasome alpha subunit 2
0.3
0.3
0.2
0.2
0.1
0.1
0.0
WT
C4
20S proteasome alpha subunit 3
0.10
0.08
0.06
0.04
0.02
0.00
WT
C4
Carbamate Kinase
0.0040
0.0030
0.0020
0.0010
0.0000
WT
C4
Hyp (ribosomal p2)
4.0
3.0
2.0
1.0
0.0
WT
C4
P21.1 (17060)
8.0
6.0
4.0
2.0
0.0
WT
C4
GTA-1
30.0
25.0
20.0
15.0
10.0
5.0
0.0
WT
C4
PPIase B precursor
25.0
20.0
15.0
10.0
5.0
0.0
WT
C4
Xaa-Pro didpeptidase
10.0
8.0
6.0
4.0
2.0
0.0
WT
C4
EF-2
3.0
2.5
2.0
1.5
1.0
0.5
0.0
WT
C4
Spindle pole protein
0.8
0.6
0.4
0.2
0.0
WT
C4
P21.1 (SMC domain - 27925)
15.0
10.0
5.0
0.0
WT
C4
TPI
2.5
2.0
1.5
1.0
0.5
0.0
WT
C4
Alpha-7.3 giardin
6.0
5.0
4.0
3.0
2.0
1.0
0.0
WT
C4

(Previous page) Figure 3.4. Differences between mRNA levels of genes corresponding to proteins differentially modulated in 2D DIGE experiments for WT and NTZr/METr *G. lamblia*. Transcript levels were measured by qRT-PCR and are shown in arbitrary units relative to beta actin (y-axis scale varies). Students t-test was applied to the data and graphs highlighted in red denote significant results ($p = <0.05$). See figure 3.5 for alternative presentation of data.

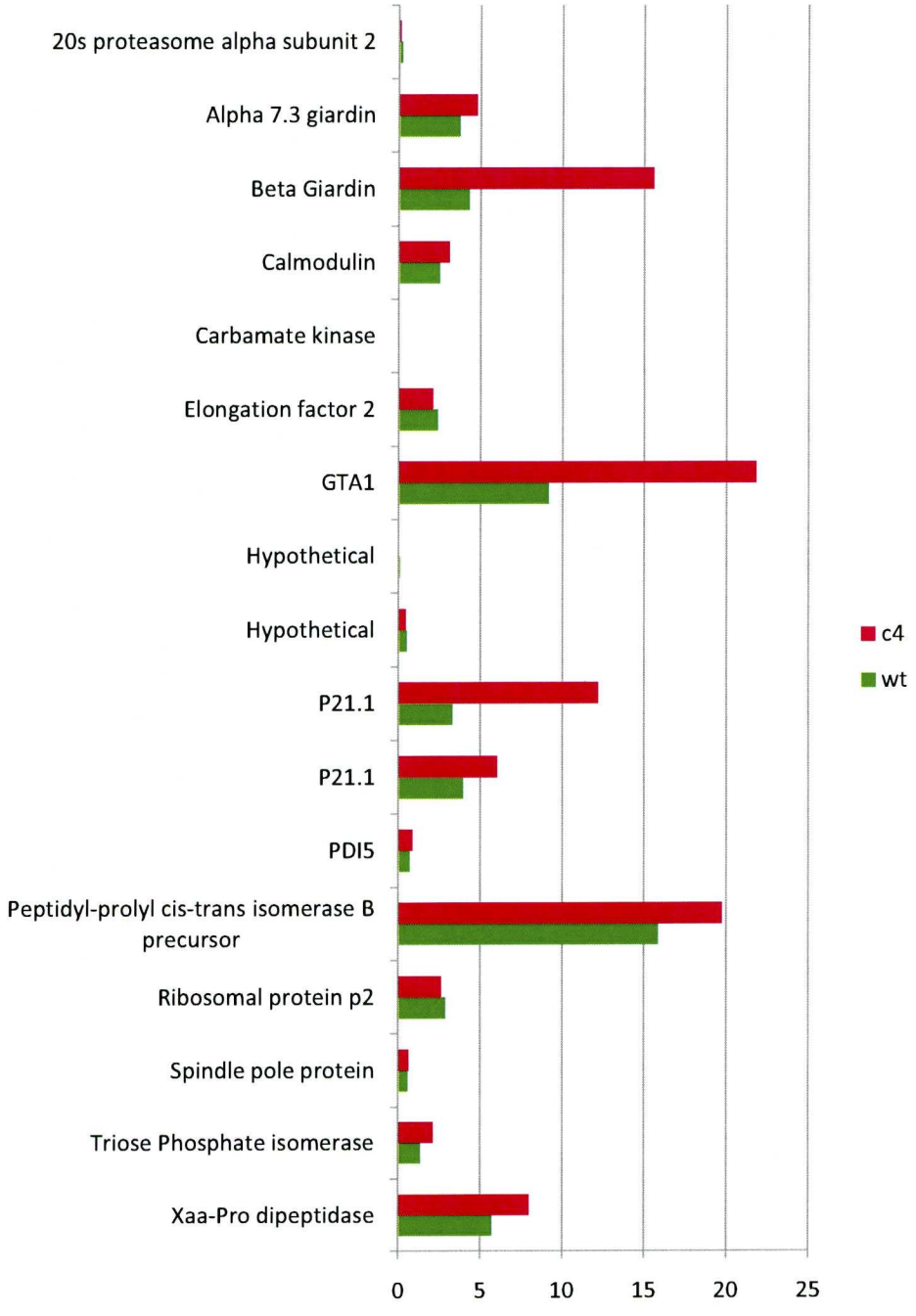


Figure 3.5. Alternative presentation of differences between mRNA levels of genes corresponding to proteins differentially modulated in 2D DIGE experiments for WT and NTZr/METr *G. lamblia*. Other details as per figure 3.4. mRNA levels displayed on consistent axis to illustrate relative abundace between genes.

Microarray re-analysis of WT vs NTZR/METR microarray data.

Two microarray screens of the transcriptome of WT *vs.* NTZr/METr *G. lamblia* were analysed using limma and limmaGUI. From the raw red/green background subtracted expression intensities M (intensity ratio) and A (average intensity) were calculated, equations for which are shown below.

$$M = \log_2 R - \log_2 G$$

$$A = \frac{1}{2} \times (\log_2 R + \log_2 G)$$

An MA plot (Figure 3.6) revealed no abnormally scattered groups of spots that might indicate problems in hybridization or damage to the array and only very slight intensity dependent effects in the data as characterised by an asymmetric "fan-tail" distribution. However the asymmetric distribution appeared to be an effect only on the empty array spots and was corrected by subsequent normalisation. However, plotting the same data according to their 'print-tip group' i.e. the printer tip with which each tip was printed, revealed considerable systematic biases in the data (Figure 3.7 (top panel)). The data were therefore normalised within arrays using the print-tip LOWESS (locally weighted scatterplot smoothing) method the results of this are illustrated in Figure 3.7 panel a (top - before) and b (bottom - after). Data did not appear to require normalisation between arrays (Figure 3.8) and had a very high correlation coefficient of >0.99. After averaging within array replicate spots, and application of a linear regression model, data were assessed using empirical Bayesian statistics, and ranked by the topTable function in order of significance

of their evidence for differential expression as measured by their 'log odds', or 'B statistic'. The resulting dataset for all genes was filtered using several methods. Empty, reserved and control spots and those

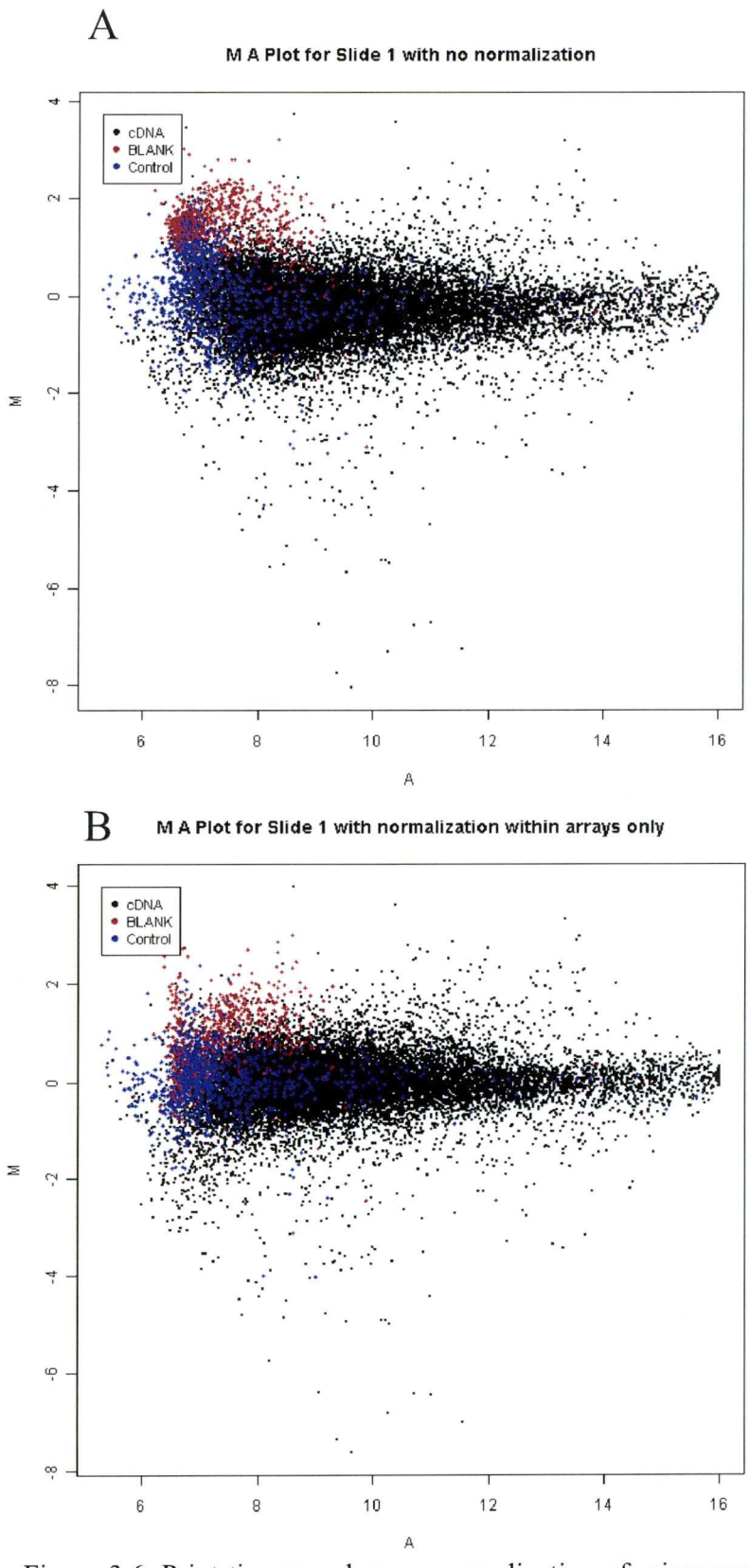


Figure 3.6. Print-tip group lowess normalization of microarray data. M A plots are shown for slide 1 (A) before and (B) after within array lowess normalization. Black spots = *Giardia* genes, Blue = *Arabidopsis* control spots, Red = Blank

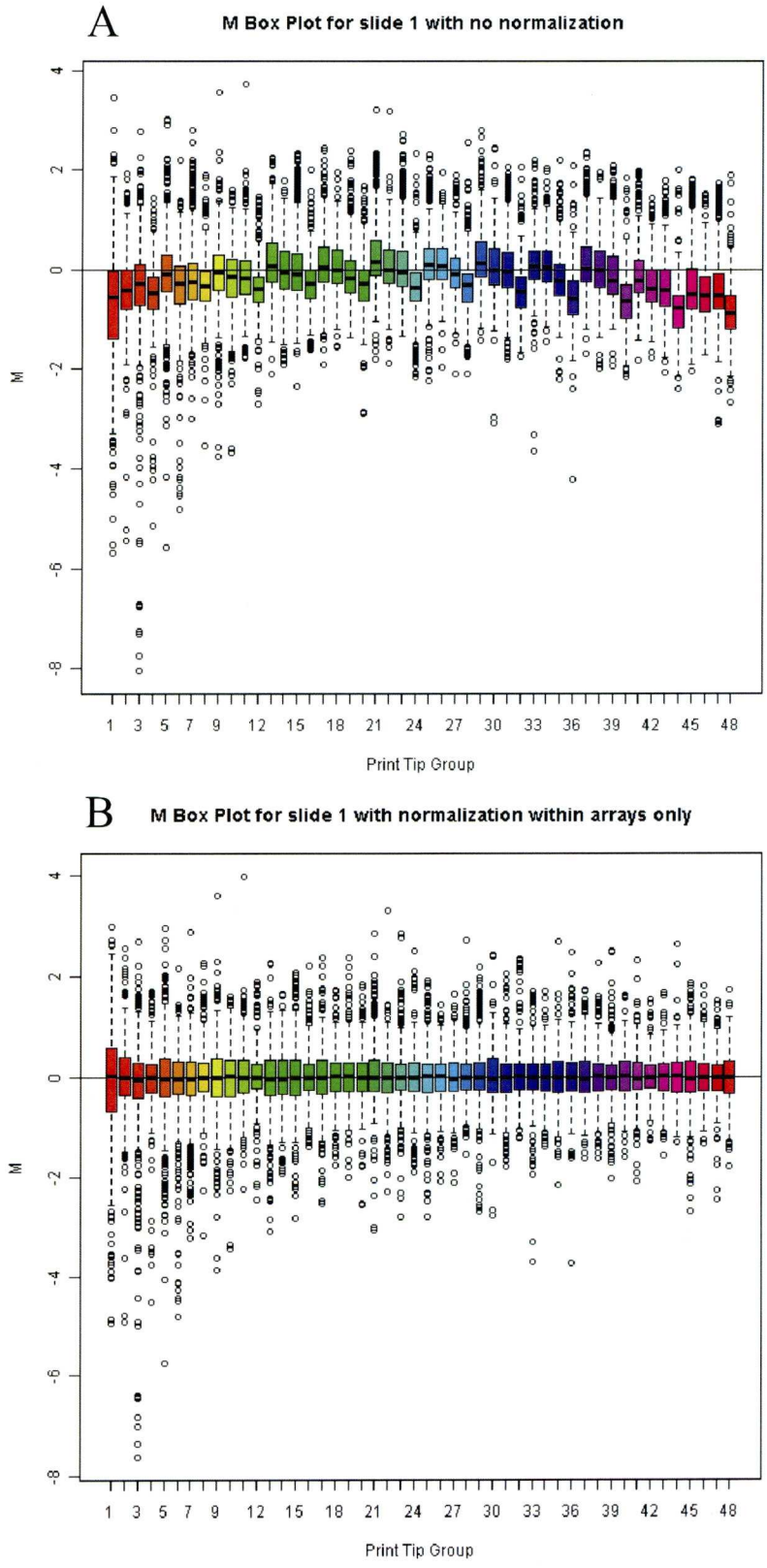


Figure 3.7. Print-tip group lowess normalization of microarray data. M box plots are shown for each print-tip group before (A) and after (B) lowess normalization.

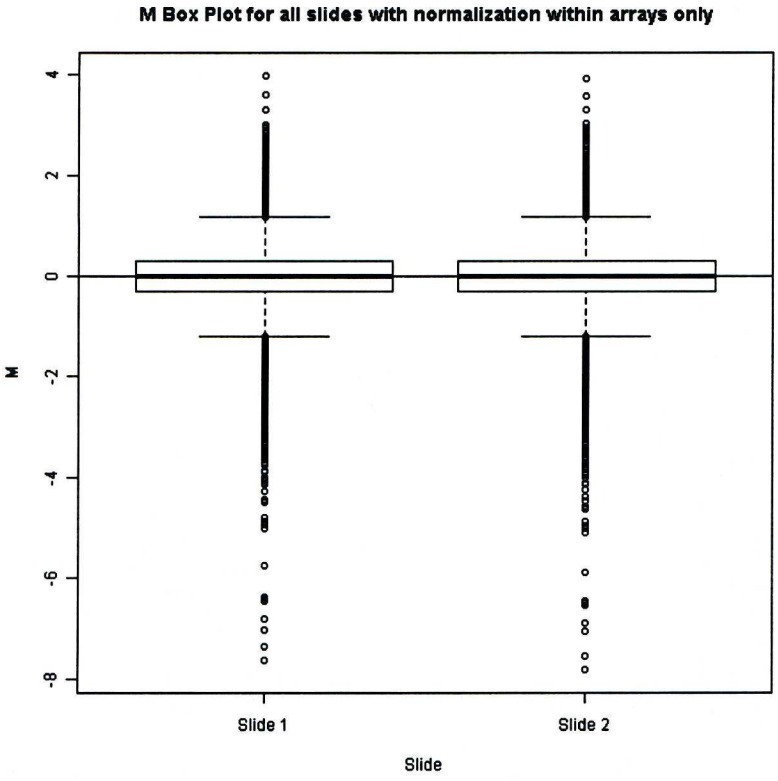


Figure 3.8. M box plots of microarray data. M box plots are shown for each microarray slide after lowess normalization..

with very low average expression (less than 500 on a linear scale) were excluded, as were genes with a $\log_2$ fold change between 1 and -1 (corresponding to a 2-fold change). Genes with Holm adjusted p-values <0.05 (a multiple testing correction) were also eliminated. Due to the relative age of the release of the gene model that was used to construct the microarray, the remaining data were cross-referenced with G. lambliaDB and depreciated genes were also eliminated. This resulted in a final subset of 278 genes including 37 that had appeared in the previously published analysis of this data . These do not include hypothetical proteins, which were not specified in the publication (see Table 3.8 and appendix for full list) and includes several large families or groups of proteins, the largest of which are summarised in Figure 3.9.

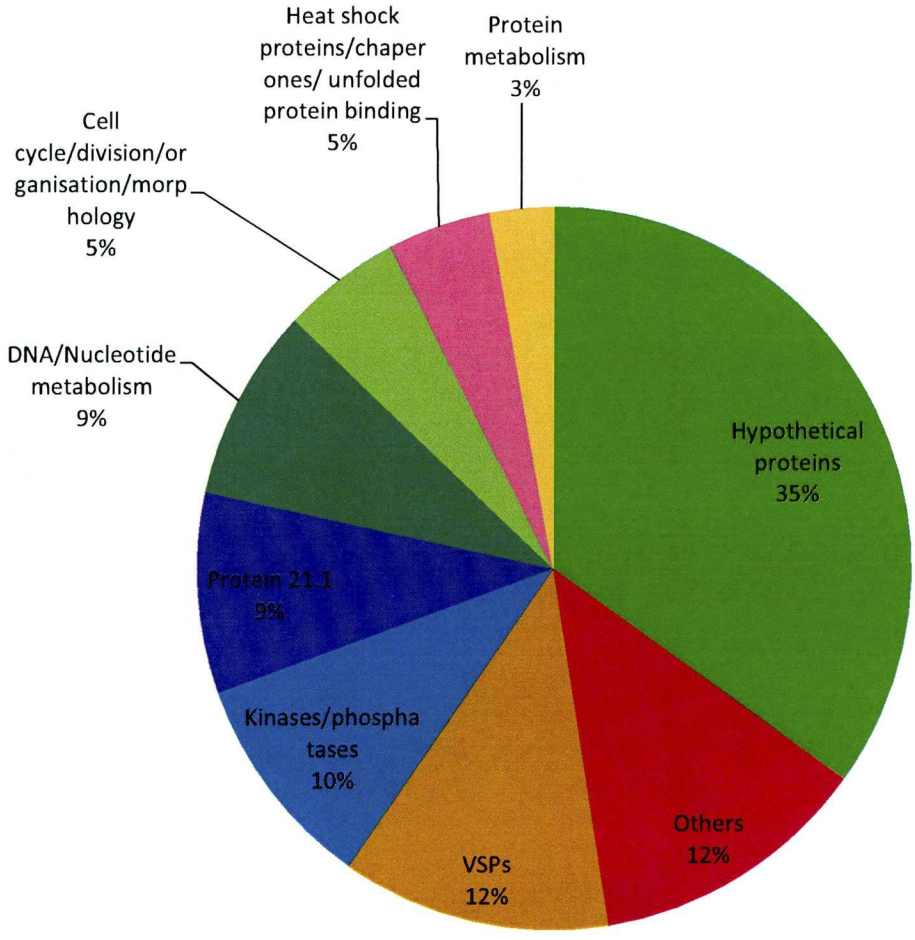


Figure 3.9 Distribution of categories of genes differentially expressed in NTZr/METr *Giardia lamblia* as determined by two colour microarray

Database ID	Product Description	Log_2 FC
13864	Heat shock protein HSP 90-alpha	2.313765
98054	Heat shock protein HSP 90-alpha	2.240585
21942	NADP-specific glutamate dehydrogenase	-1.62711
103676	Alpha-tubulin	1.494247
5435	Cyst wall protein 2	-1.45207
17249	Coiled-coil protein	1.48684
10429	Wos2 protein	2.277021
27310	Stress-induced-phosphoprotein 1	1.419363
115066	High cysteine membrane protein VSP-like	1.378404
7260	Aldose reductase	1.116979
7031	Spindle pole protein, putative	1.793511
4812	Beta-giardin	2.118459
12139	Protein 21.1	2.238835
8722	Myb 1-like protein	-1.13812
137672	High cysteine membrane protein VSP-like	-1.22458
27925	Protein 21.1	2.448497
112103	Arginine deiminase	1.654431
10193	WD-repeat membrane protein, putative	-1.27463
15411	Kinase, NEK-frag	1.121846
17043	Glyceraldehyde 3-phosphate dehydrogenase	-1.01292
15953	Kinase, NEK	1.060772
16745	Axoneme-associated protein GASP-180	1.253151
8174	Protein 21.1	1.412708
5188	Protein 21.1	1.673959
16967	Kinase, NEK	1.200261
15026	Src-associated protein-like protein	1.159118
17327	Xaa-Pro dipeptidase	1.075094
24590	Protein 21.1	1.09809
6140	Kinase, NEK	-1.26339
8173	Glycerol kinase	-1.45102
16125	Glycerol-3-phosphate dehydrogenase	-1.23923
8726	Actin related protein	1.61917
16906	Phosphatidate cytidylyltransferase	1.084544
5638	Cyst wall protein 1	-1.10152
16124	TCP-1 chaperonin subunit eta	1.267606
15228	Ribosomal protein S15A	-1.18637
17551	Protein 21.1	2.238335
101291	Beta tubulin	1.32733
17570	Elongation factor 2	1.097469
4410	SALP-1	1.30995
16283	Amino acid transporter family	-1.4905
11992	TCP-1 chaperonin subunit epsilon	1.751945
21444	Spindle pole protein, putative	1.679051
15455	E04F6.2 like protein	1.067718
16532	Protein 21.1	2.38583
137726	ABC transporter ABCA.1, putative	-1.13739

9270	Dual specificity protein phosphatase CDC14A	1.528975
17053	Protein 21.1	1.884784
88369	Protein 21.1	1.295851
10255	Translation initiation factor eIF-4A, putative	1.107429
16412	Heat-shock protein, putative	1.087815
113622	Protein 21.1	1.759351
8803	Protein 21.1	-1.09839
16867	AAA family ATPase	1.466503
9515	Coiled-coil protein	1.35918
10219	Protein 21.1	1.080224
88765	Cytosolic HSP70	2.587132
7103	Kinase, NEK	-1.11853
96460	Alanyl-tRNA synthetase	1.629698
10570	FKBP-type peptidyl-prolyl cis-trans isomerase	1.508949
135231	Histone H3	1.035544
5489	Kinase, NEK-frag	1.001745
100955	Mitotic spindle checkpoint protein MAD2	1.686949
9046	Sugar transport family protein	-1.15105
23492	Protein 21.1	1.365797
1695	Rab11	1.022995
16453	Carbamate kinase	1.999925
8445	Kinase, NEK	1.288165
3762	Protein 21.1	1.146573
93011	Protein 21.1	1.326728
17097	Protein 21.1	1.032724
17587	CTP synthase	1.455262
9720	Protein 21.1	1.168541
16235	Kinase, CAMK CAMKL	1.320865
14628	Phosphatidylinositol-4-phosphate 5-kinase, putative	1.002283
16272	Kinase, NEK	1.010616
8044	Seven transmembrane protein 1	1.168636
17046	Protein 21.1	2.050509
14225	CXC-rich protein	-1.00507
98126	High cysteine protein	-1.40177
7439	Ser/Thr phosphatase 2A, 65kDa reg sub A	1.694196
10370	ATP/GTP binding protein, putative	1.074461
15410	Ser/Thr protein kinase	1.222071
14568	Ser/Thr protein phosphatase PP1-alpha 2 catalytic subunit	1.290058
16839	Kinase, NEK-frag	1.017439
16376	ATP-dependent RNA helicase p47, putative	1.047976
3643	70 kDa peptidylprolyl isomerase, putative	1.625197
103807	Protein 21.1	1.049963
17060	Protein 21.1	1.837992
17558	Kinase, CMGC DYRK	1.458509
9062	Long chain fatty acid CoA ligase 5	1.056904
14859	Protein 21.1	2.074651
17585	Protein 21.1	1.316811

9548	Cathepsin L precursor	-1.51624
40390	Protein 21.1	1.096409
15048	ATP-dependent RNA helicase-like protein	1.560042
86676	Delta giardin	1.228856
17571	Trichohyalin	1.33616
11309	High cysteine membrane protein Group 1	1.665564
11165	Protein 21.1	1.053181
6199	Kinase, STE Dicty2	-1.08261
8855	P115, putative	1.024114
7268	Protein 21.1	1.413642
1657	Translation initiation factor	-1.10527
6242	Translationally controlled tumor protein-like protein	1.191121
15823	1,4-alpha-glucan branching enzyme	1.220688
14553	CCAAT-binding transcription factor subunit C	-1.24779
41288	Mucin-like protein	1.001736
11204	Plant adhesion molecule 1	-1.41588
40376	High cysteine non-variant cyst protein	-1.02141
33689	Dipeptidyl-peptidase III	1.009179
14578	Kinase	-1.22841
8364	Thymidine kinase	2.183257
16205	Kinase, NEK	-1.00998
17230	Gamma giardin	2.238187
17120	CEGP1 protein	-1.16244
4355	Nif3-related protein	-1.16136
11247	Ribosomal protein L13a	1.033089
2661	Cyclin-dependent kinases regulatory subunit	-2.4402
16001	Glutamate-cysteine ligase	-1.57894
102051	Amino acid transporter family	-2.09489
16569	Transcriptional regulator, Sir2 family	-1.75134
17578	Kinase, NEK-frag	1.330388
113021	Acetyl-CoA carboxylase/pyruvate carboxylase fusion protein, putative	-1.12655
10361	P60 katanin	-1.43448
15344	MCM2	-1.07227
7648	SMC6 protein	-1.09617
13464	Splicesome-associated protein	-1.48997
23496	RNA polymerase AI large subunit	1.255903
14821	HesB domain-containing protein	-1.27966
6602	Protein 21.1	-1.54236
12164	Small nuclear ribonucleoprotein Sm D3, putative	-1.06471
13964	Kinase, NEK	-1.12532
103164	H-SHIPPO 1	1.536103
17188	Kinase, NEK	-1.96401
10666	Histone acetyltransferase GCN5	-1.20642
33083	Mlh2-like protein	-1.1201
11953	Coatomer alpha subunit	-1.9776
114773	Cysteine protease	1.15101
112866	Reverse transcriptase/endonuclease, putative	1.190626

8152	Kinase, NEK	-1.68594
17495	RAD50 DNA repair protein, putative	-2.01367
16841	DinF protein	-1.10705
1894	U3 small nucleolar RNA-associated protein 11, putative	1.06382
113610	GlcNAc-PI synthesis protein	-1.55766
15248	Spindle protein, putative	-1.27922
5013	Endothelin-converting enzyme 2	-1.6208
14872	Protein 21.1	-2.62251
35428	Valine-tRNA ligase	-1.09424
6404	Kinesin-9	-2.80198

Table 3.8. List from "topTable" of most differentially expressed genes excluding genes correcsponding to hypothetical proteins and VSPs. Genes listed exhibited Log_2 ratio >1.0 or <-1, Holm adjusted p-value ≤ 0.01, and b-statistic corresponding to a $\geq$99 % chance of differential expression. The order of the genes in the table is based on the significance of the evidence for their differential expression (B-statistic). A full list is available in the appendix.

Bioinformatic analyses

Bioinformatic analysis of proteomic data

To describe the proteomic data in a more systematic manner, and to enable the comparison of broader areas of the biology of *G. lamblia* without reference to specific proteins and genes, GO terms were also used to describe the data from the DIGE experiments. Hypergeometric distribution analysis is commonly applied to this type of hierarchical data, however proper statistical tests cannot be applied to these data due to differences between proteomic and genomic 'background populations'. Unlike genomic experiments, the pool from which proteins are picked may differ significantly from the theoretical pool required to make statistical calculations. This can be the result of things such as cell differentiation or life-cycle stage. Furthermore, the sample size is insufficiently large for hypergeometric analysis. The GO terms are still useful in describing in a simple and logical manner the types of proteins in each group however

GO annotations occurring in both DIGE experiments

To identify GO terms that might be present due to sub-lethal doses of drug in their growth media, Lists of individual GO terms from both the WT vs. NTZr/METr and WT vs. NTZr/METr + NTZs/METr DIGE experiments were compared and terms appearing in both DIGE experiments were compiled. Since assigning the most specific GO term possible also implicates every term preceding it in the hierarchy by default, the very small number of proteins used creates a relatively large network of GO terms that are best illustrated in their

native hierarchal layout (Figure 3.10). The small number of proteins in this group however means that little information outside of that known about the individual proteins can be gleaned from this data.

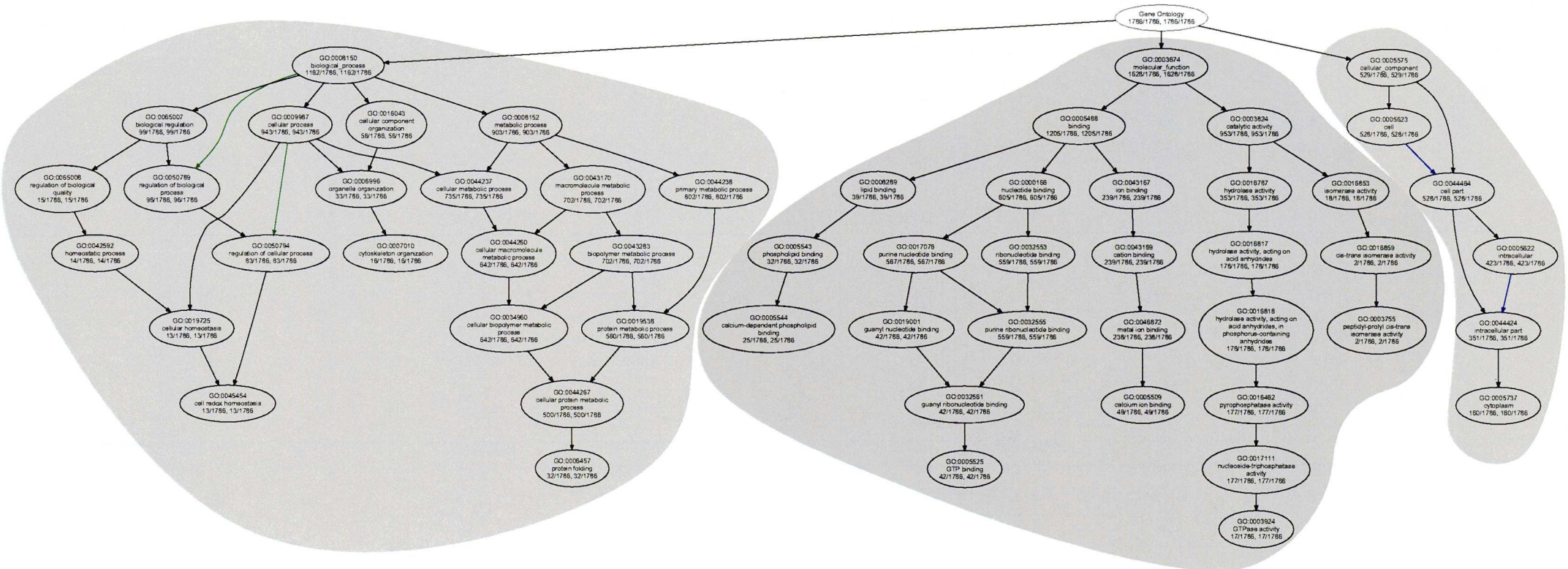


Figure 3.10. Full lineages of all GO terms associated with changes occuring in both WT vs. NTZr/METr and WT vs. NTZr/METr + NTZs/METr DIGE experiments.

GO annotations only in WT vs. NTZR/METR DIGE

Lists of individual GO terms from both the WT vs. NTZr/METr and WT vs. NTZr/METr+NTZs/METr DIGE experiments were compared and terms appearing uniquely in the WT vs. NTZr/METr experiment were compiled. The larger number of proteins in this group suggests that there are potentially significant differences between the specific responses of the NTZr/METr parasites and a generic drug or stress response. The greater number of proteins also means that a greater number of GO terms are implicated. The terms and full lineages are shown in their hierarchal layout in figures 11 (biological process),12 (molecular function) and 13 (cellular component). In summary, the biological processes of translation, protein synthesis and proteolysis were heavily present. The nuclear pore and ribosomes were highlighted as being important cellular components. The molecular functions implicated are more difficult to summarise due to a smaller number of available annotations, although catalytic functions, particularly peptidase and isomerise activity were implicated, as were (metal) ion binding, and the activity of structural molecules.

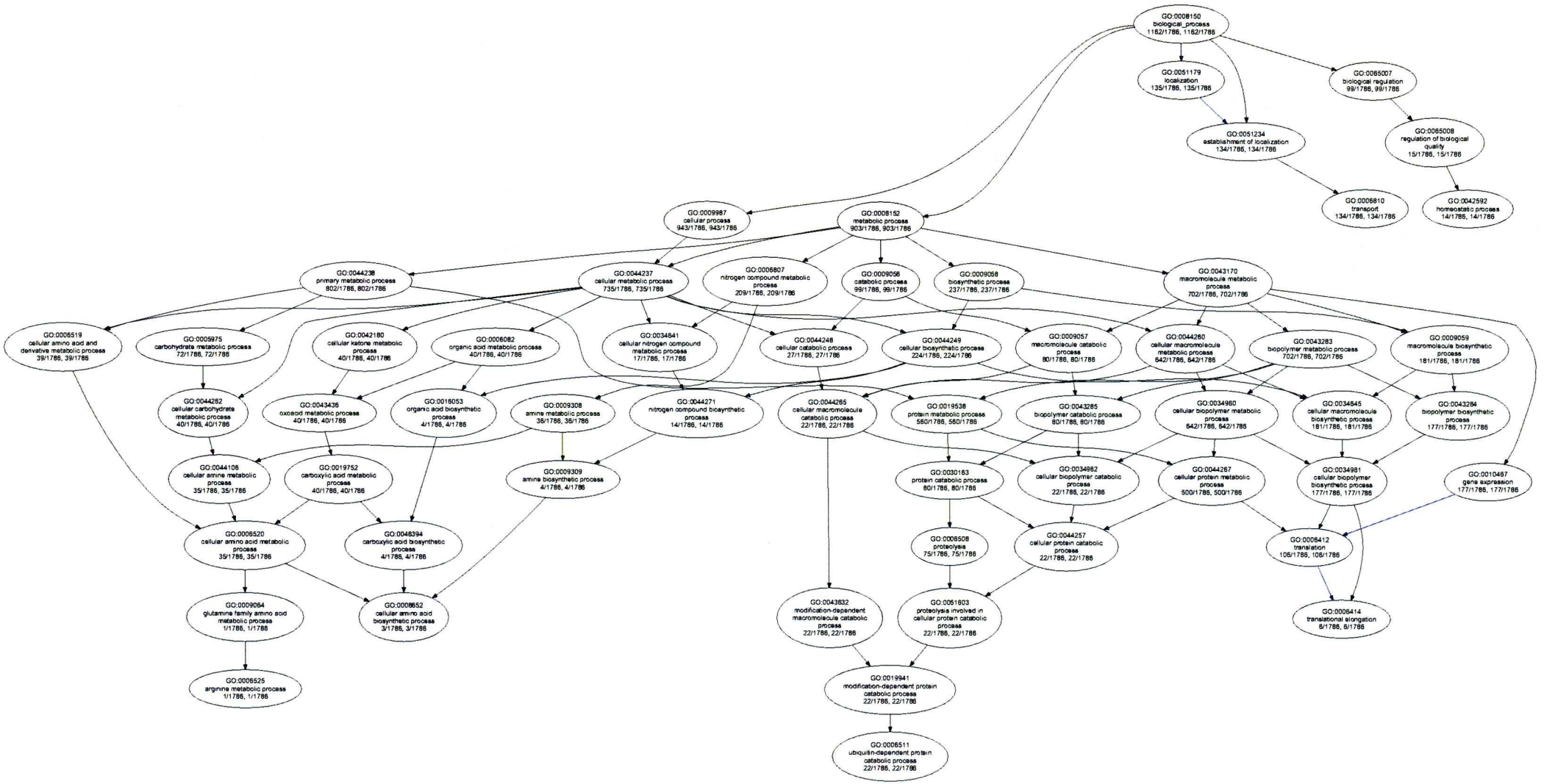


Figure 3.11. Full lineages of "Biological process" GO terms associated with solely with changes detected in WT vs. NTZr/METr DIGE experiments.

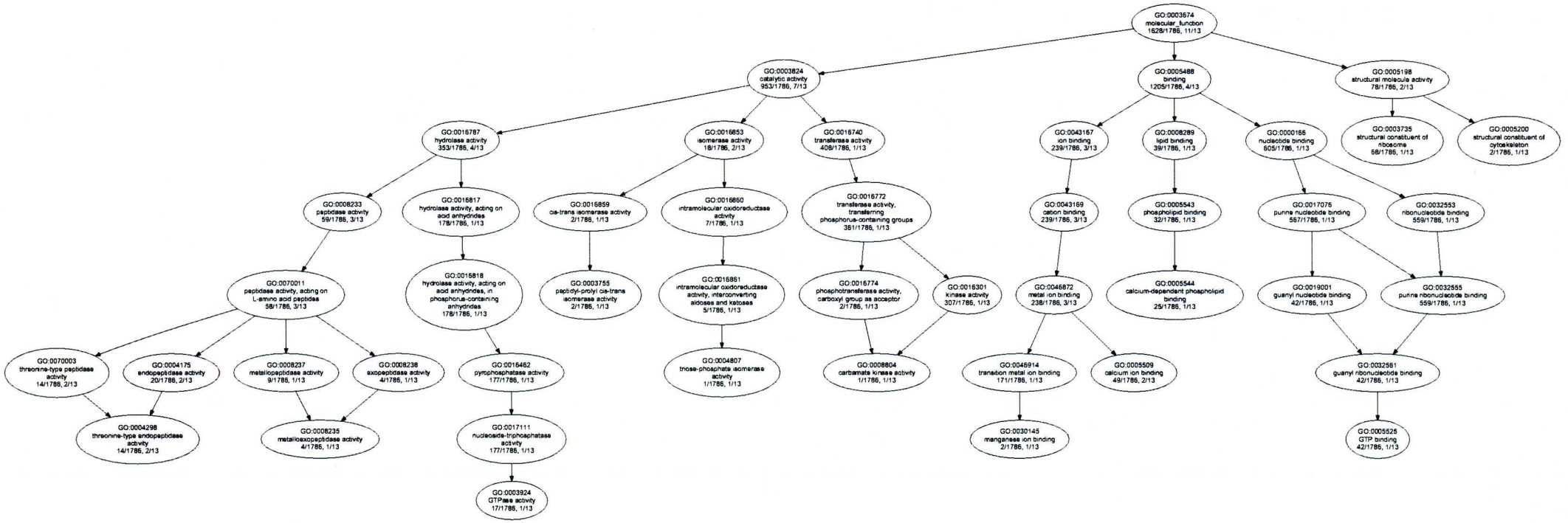


Figure 3.12. Full lineages of "Molecular function" GO terms associated with solely with changes detected in WT vs. NTZr/METr DIGE experiments.

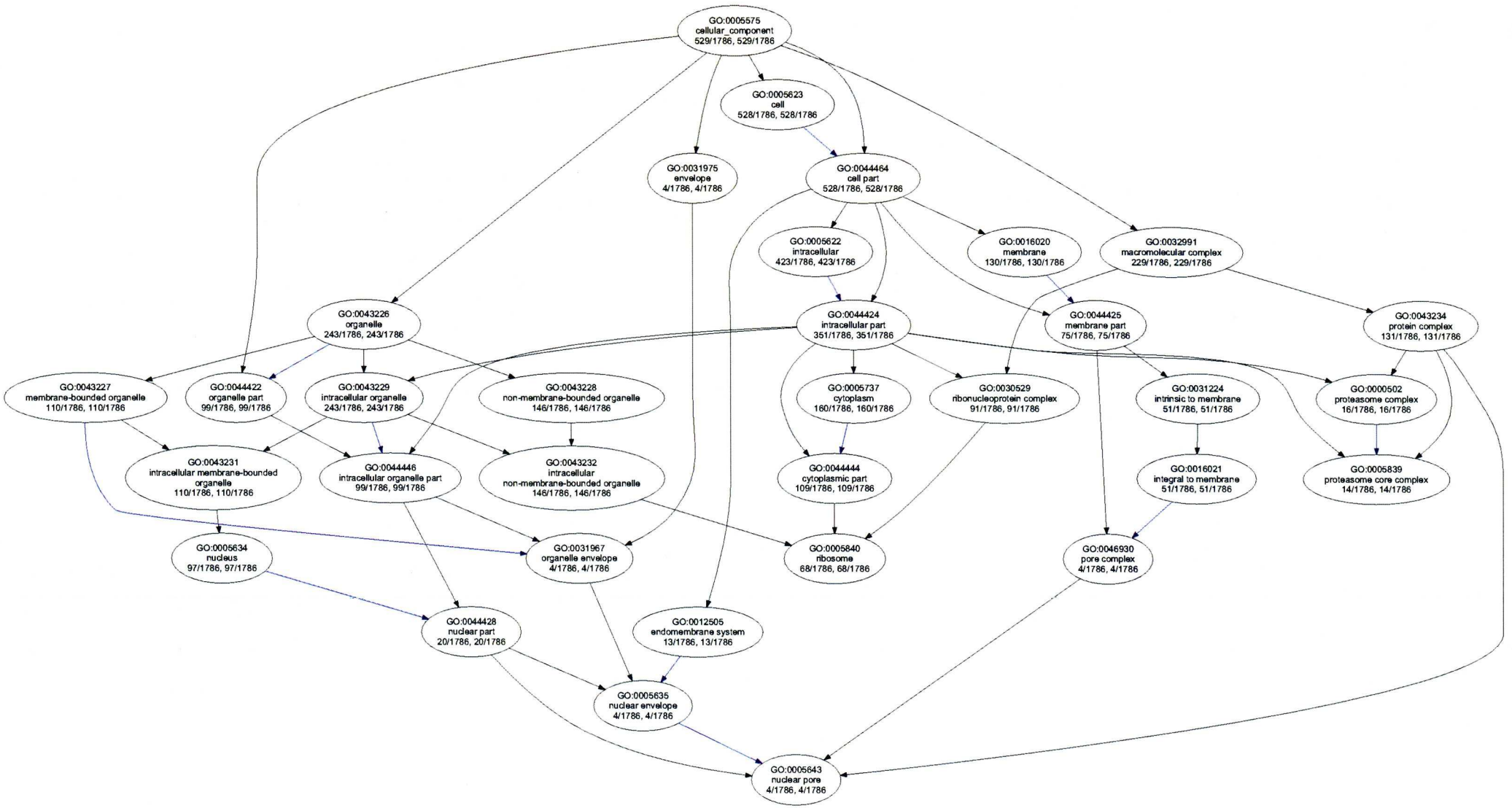


Figure 3.13. Full lineages of "Cellular component" GO terms associated with solely with changes detected in WT vs. NTZr/METr DIGE experiments.

Bioinformatic analysis of microarray data

Hypergeometric distribution analysis of GO terms

Using the filtered topTable list of differentially expressed genes given in Table 2.9, Ontologizer was used to ascertain which (if any) GO terms were overrepresented using hypergeometric distribution and the Parent-child method (Figure 3.14). Significantly ($p \leq 0.05$) overrepresented terms are highlighted in colour with greater intensity of the colour denoting greater significance. The most significant GO terms were unfolded protein binding and protein folding. The other significant terms are also given in tabular form in Table 3.9.

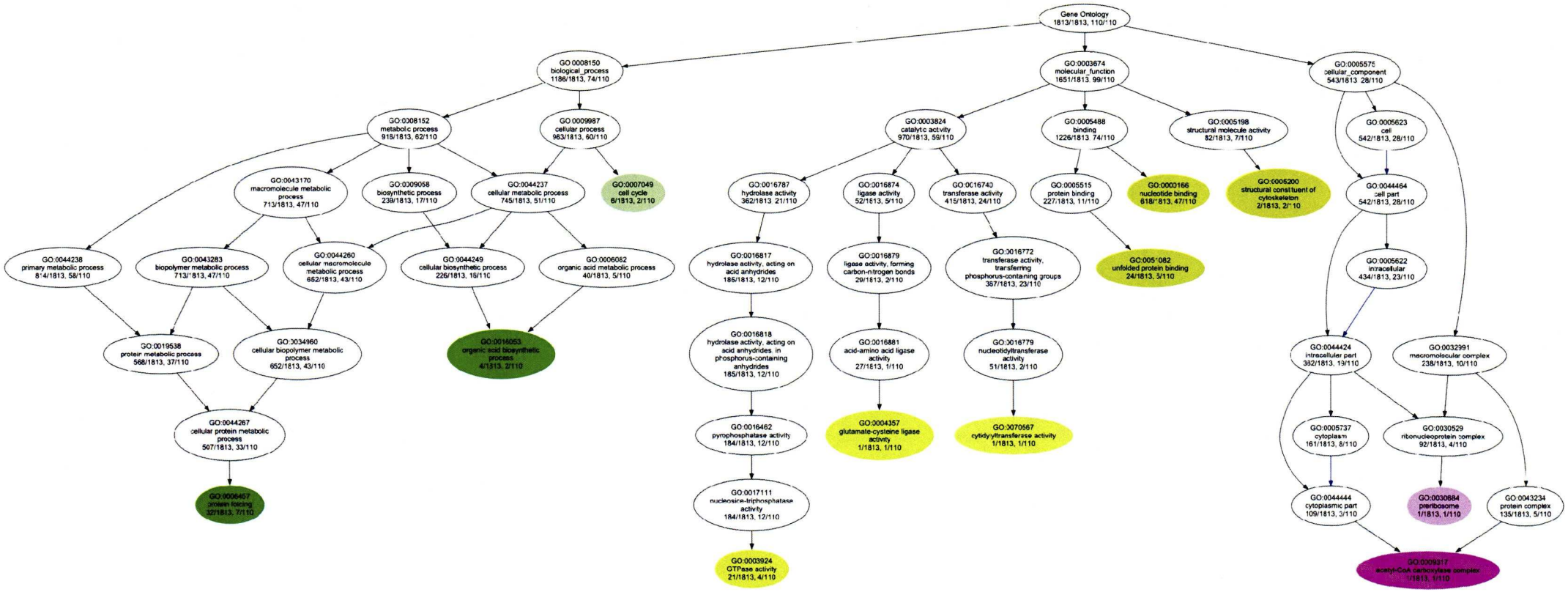


Figure 3.14. GraphViz vizualisation of significantly overrepresented GO terms using the Parent-child method. Ontologizer 2.0 was used to determine which GO terms were significantly overrepresented in the topTable list of genes determined to be differentially expressed in the microarray experiment.
Colours represent GO terms emanating from: Green = biological process, yellow = molecular function, magenta = cellular component. Higher intensity colours signify higher significance

GO ID	Name	GO Aspect
GO:0051082	unfolded protein binding	M
GO:0006457	protein folding	B
GO:0005200	structural constituent of cytoskeleton	M
GO:0000166	nucleotide binding	M
GO:0003924	GTPase activity	M
GO:0016053	organic acid biosynthetic process	B
GO:0005874	microtubule	C
GO:0009317	acetyl-CoA carboxylase complex	C
GO:0005694	chromosome	C
GO:0004357	glutamate cysteine ligase activity	M
GO:0070567	cytidylyltransferase activity	M
GO:0030684	preribosome	C

Table 3.9. Hypergeometric distribution analysis of differentially expressed genes using the Parent-child union method. GO Aspect describes the type of term, M = Molecular function, B = Biological process, C = Cellular component.

Discussion

2D-DIGE was used to determine which proteins are modulated in a strain of *G. lamblia* resistant to NTZ (NTZr/METr) relative to a wildtype strain (WT). Using qRT-PCR, we also quantified the levels of mRNA of the genes encoding the differentially expressed proteins to gain further insight into the underlying biology of their expression. By using the results of a second DIGE experiment that also included a NTZs/METr strain of *G. lamblia* it was hoped that proteins modulated due to MET resistance, or the stress of being grown under drug treatment conditions could be identified. An existing microarray dataset comparing WT and NTZr/METr *G. lamblia* was also re-analysed to gain further insight into the transcriptomic changes occurring in this strain during the acquisition of resistance.

There are five proteins that were identified in both the WT vs. NTZr/METr experiment and the WT vs. NTZr/METr + NTZs/METr DIGE experiment. This could suggest that their presence could be due to a generic response to the sub-lethal growth conditions (Table 3.6). The proteins were identified as Protein 21.1 (P21.1) (GL50803_17060), P21.1 – containing a putative bacterial type SMC domain (GL50803_27925), Peptidyl-prolyl cis-trans isomerase, Elongation factor 2 and Alpha 7.3 giardin. Of these proteins, the level of transcript corresponding to the two P21.1s and Alpha 7.3 giardin differed significantly between the WT and NTZr/METr strain. The measured increase in abundance of Alpha 7.3 giardin protein corresponded well to the observed increase in transcription, however the regulation of the two P21.1s, appears to be more complex. For P21.1 (GL50803_17060), proteins spots that

were both up- and downregulated were detected in both DIGE experiments. Plotting their observed migration distance against their hypothetical mass (from G. lambliaDB) (Figure 3.3) shows that although one of the spots for this protein appears in its expected location (GL50803_17060), the two other instances appear at a much lower mass of approximately 20-25 kDa. Using "Compute pI/Mw" tool at ExPASy [117] to predict just the portion of the predicted protein that contains peptide hits, yields a Mw of ~22,032 Da. This portion corresponds with a predicted ankyrin repeat domain and is consistent with the hypothesis that this is the result of either proteolysis or alternative splicing. Little information about this P21.1 (GL50803_17060) is currently available.

Although the other P21.1 (GL50803_27925) was also present in both DIGE experiments, closer inspection of the results reveals that this protein was downregulated in one experiment and upregulated in the other (unlike the other proteins that were all modulated in the same direction) and thus may still be important in NTZ resistance. The increase in transcript levels between WT and NTZr/METr strains for the corresponding gene, as measured by qRT-PCR, is one of the largest measured here and was also detected in the re-analysis of the microarray data with a similar fold increase. Both P21.1 (GL50803_27925) and a Hypothetical protein (GL50803_15499 – protein upregulated, transcript not measured) are predicted to possess bacterial-type SMC (structural maintenance of chromosomes) domains. SMC containing proteins are part of the condensin complex that plays a central role in chromosome assembly and segregation. Chromatin condensation, spindle

microtubule nucleation and the alignment of the nuclei along the dorsal–ventral axis, all occur during early mitosis in *G. lamblia* (a binucleate organism). The opposite directions of modulation of this P21.1(27925) in the two DIGE experiments, alongside the identification of a hypothetical protein containing a similar SMC domain and Spindle pole protein being modulated in NTZr *G. lamblia*, suggests that cell cycle events and mitosis may be significant events in NTZ resistance.

Other proteins with cell cycle and mitosis related functions were also found solely in the WT vs. NTZr/METr DIGE; Calmodulin (upregulated), Hypothetical nucleoporin autopeptidase (upregulated), and Beta giardin (protein downregulated, increased mRNA) and a possible Dyenin light chain (downregulated). Beta-giardin localizes to the ventral adhesive disk of *G. lamblia* and appears to be homologous with SF-assemblin. It has been proposed that SF-assemblin is involved in reconstructing the microtubular structures in the cell after mitosis [118, 119]. Structural observations of *G. lamblia* cells undergoing mitosis have pointed to the ventral disk being an important novel component of nuclear division [120, 121] supporting the hypothesis that beta giardin is involved in this process. The decreased level of beta giardin in the NTZ resistant strain is also consistent with the findings of Muller *et al* [106] who also observed significant down regulation of several genes involved in cell division in a comparison of the same cell types. Also included in this group is a hitherto hypothetical protein that appears to be a nucleoporin autopeptidase, part of the nucleoporin complex. Regulation of the nucleoporins is believed to contribute to the control of the cycle [122] and

different components of the nuclear pore complex (NPC) regulate differential gene expression dependent on the current phase in the cell cycle. We have observed modulation of this protein but with no significant change in transcription. This is similar to that observed in HeLa cells by other markers, where levels of Nup96 (also a nucleoporin) are controlled by ubiquitin mediated proteasomal degradation rather than new transcription [122]. Modulation of 2 components of the 20s proteosome that are believed to be involved in this process were modulated. In the case of Nup96, increased degradation leads to an increased cell cycle speed and *vice versa* [122]. The closest homologue appears to be Nup98 however, a protein often implicated in several types of cancer [123], which has functions in transport, viral infections and mitosis [122].

Changes in proteins involved in cell cycle events and chromosome organisation could be indicative of a cell that is preparing to undergo mitosis [124], or perhaps enter another life cycle stage. i.e. encyst. This may be a possible explanation for the observed cell cycle protein changes and changes in proteins related to organisation of the chromosome. If resistance does involve proteins involved cell cycle progression we might also expect to see differences in growth rates between NTZr/METr, NTZs/METr, and WT cell lines, and indeed the NTZs/METr strain grows more slowly than the NTZr/METr clone which in turn grows more slowly than the wildtype strain, potentially indicating that the resistant strains have alterations in their growth mechanisms [28].

In further support of cell cycle events being important factors in NTZ resistance, the microarray data reveals that a great many VSPs are differentially expressed in the NTZr/METr strain. Svard *et al* [125] previously noted widespread changes in VSPs upon encystation, particularly downregulation and internalisation of the main surface protein, TSA417, and upregulation of other VSPs, which is presumed to be a host immune evasion response. However, TSA417, which was ranked highest in the microarray 'topTable', with levels of transcript downregulated approximately 10 fold (-3.4 log_2FC) in the NTZr/METr strain, was also observed as being over 5 fold downregulated in the WT vs. NTZr/METr+NTZs/METr DIGE experiment, suggesting that part of this process could be a common response to NTZ and MET.

Another notable group of proteins modulated solely in the NTZr/METr strain are those of protein degradation:- 20s proteasome alpha subunit 2 (upregulated), 20s proteasome alpha subunit 3 (downregulated) and Xaa-Pro dipeptidase (upregulated). Changes in the levels of proteins responsible for wider scale protein degradation might indicate a need for the cell to breakdown proteins that have either been misfolded or are unfolded. The 20s core of the proteasome is responsible for a number of protein degradation functions including the degradation of aggregated oxidised proteins [126, 127]. The increase in chaperone proteins (particularly HSP90) in the microarray experiments also supports the notion that NTZ might cause an increase in misfolded proteins. In this scenario, NTZ resistance could be brought about by

an increase in protective enzymes that repair proteins that would otherwise be lethal to the cell.

In addition to the increases in chaperone type proteins, the microarray data showed significant changes in the expression of several large groups of genes (Figure 3.9 and Table 3.8). The VSPs and P21.1s are particularly prominent examples of this. These VSP changes might be indicative of epigenetic changes occurring in response to drug treatment that are normally deployed in *G. lamblia* to counter host immune defences [128]. Epigenetic changes in gene expression are not regulated by changes in the underlying genomic sequence but by other mechanisms such as DNA methylation or histone acetylation which can 'activate' whole regions of DNA. Widespread constitutive changes in expression of this type might also mask a change in a protein(s) that could be directly responsible for resistance to NTZ however. The fact that resistance to NTZ and MET is maintained in the absence of NTZ during trophozoite growth but easily lost after a round of en- and excystation [106] further supports the notion that it is an epigenetic change in gene/protein expression, rather than a genetic mutation that is leading to resistance.

Since biochemical pathway analysis is still a field in its infancy, and interaction information for proteins in *G. lamblia* is virtually non-existent, a more abstracted approach to identifying areas of biological significance was also taken. GO terms were used to describe the data in broader terms, and has highlighted some areas for investigation that might not have been observed by considering proteins in isolation. There are some difficulties applying this technique to *G. lamblia,* since annotation of the genome with GO terms is still

largely limited to electronic annotation from homologous proteins and structural motifs. This will develop over time and, with further *G. lamblia* specific annotation, may yet still reveal more interesting biology other than what will be discussed here. Furthermore, whilst the information was freely available, this task required the manual extraction and compilation of data into formats readable by the software used.

From the proteomic data, there were several GO term lineages that appear solely in the NTZr/METr strain that warrant further study. In terms of biological processes it appears that the synthesis of amino acids, the metabolism of proteins and translational elongation appear to be important (Figures 11, 12), as do the ribosome and nucleus (nuclear pore) in terms cellular locations (Figure 3.13). This suggests that changes in protein synthesis are potentially very important in NTZ resistance. Extending this type of analysis to the microarray data which is better suited to this type of analysis, allows us to look for statistically overrepresented terms using hypergeometric distribution. Table 3.9 and Figure 3.14 show the that analysis of the 'topTable' genes from the microarray experiment indicates that the most significantly overrepresented terms are Protein folding and Unfolded protein binding and also cytoskeletal components may be important. The latter result changes depending on the method used for analysis however the "parent-child" analysis method used was chosen as a result of its more conservative approach compared to the other two common methods. The identification of Protein folding, Unfolded protein binding and cytoskeletal components as important, appears to support the results from the proteomic data and although, there are

not very large overlaps in the actual genes from the proteomic and transcriptomic data, by considering the results using GO terms, we can see that highly related processes appear to be occurring.

Summary

Using data from both proteomic and transcriptomic sources and using multiple drug resistant strains of *Giardia lamblia,* information about possible genes, proteins and biological mechanisms involved in resistance to NTZ have been investigated. Cross resistance of the NTZr stain used presented a particular problem. This was addressed with the inclusion of a NTZs/METr strain in the analysis. This approach has highlighted some interesting proteins and biological processes that warrant further study. Control of cell cycle events appears to be particularly important, with many of the differentially expressed proteins possesing functions relating the cell cycle, mitosis, and chromosomal organisation. The proteomics experiments also indicated that protein degradation might also be important, and this was further implied by the bioinformatic analysis of the microarray data which, by using gene ontologies (GO terms) to describe the data, has specifically implicated the areas of amino acid synthesis, metabolism of proteins, translational elongation, protein folding, unfolded protein binding and components of the cytoskeleton as being of importance. None of the changes observed in these experiments suggested that modulation of the expression or function of PFOR is of particular significance to these NTZr strains however.

Using a global approach to study resistance to NTZ in an organism with immature genome annotation has imposed some limitations in interpreting the results (many uncharacterised hypothetical proteins for example). Revisiting the data in time, might still reveal useful information regarding the mechanism of

resistance to NTZ in *G. lamblia*. Meanwhile, proteomic experiments in human hepatocellular carcinoma cells expressing hepatitis C virus proteins have been performed (Chapter 4). Human proteins and their interactions are better characterised and the results should result in fewer functional annotation limitations.

Chapter 4

Modulation of Hepatic Replicon (RP7) Cell Proteome in Response to Treatment with NTZ

Introduction

Knowledge of the anti-viral properties of the thiazolides is increasingly widely publicised and thiazolides are believed to be a viable option for improving the treatment of hepatitis C virus (HCV) [9-11, 13, 14, 16-18, 129]. Used as a monotherapy, data on the efficacy of NTZ suggests it is of limited use for the treatment of HCV. NTZ is currently in advanced stages of clinical trials for the treatment of HCV when used in combination with other existing standard of care (SOC) therapies such as pegylated interferon and ribavirin however, since in combination with these existing treatments, sustained virologic responses (SVR) are observed [130].

The anti-viral properties of the thiazolides raise new questions regarding the mode of action of NTZ that cannot be satisfactorily addressed by previous research from non-viral targets. One of the main appeals of the thiazolides as a treatment for Giardiasis and anaerobic bacterial infections was the widely held belief that it elicited its efficacy via targeting of PFOR via its cofactor; as such it would be both specific to organisms that relied on anaerobic energy metabolism to survive, and difficult for resistance to be acquired. Viruses do not possess PFOR in their genomes and hence the mode of action against them must lie elsewhere. Recent developments in eliciting the mode of action of the thiazolides against HCV has led researchers to believe that they act by directly

increasing protein kinase R (PKR) auto-phosphorylation and activation [131, 132]. This response normally occurs in the presence of double stranded RNA (dsRNA) [133], but the mechanism by which NTZ augments this process is yet to be described. Activation of PKR, is an important host anti-viral response which leads to phosphorylation of elongation initiation factor 2α (eIF2α) and subsequently reduced protein synthesis and reduced levels of unfolded protein and several other downstream effects. When effected by PKR this process is termed the integrated stress response (ISR). However, it can also be induced as part of the unfolded protein response (UPR) where protein kinase-like endoplasmic reticulum kinase (PERK) takes on the role of PKR in response to high levels of unfolded protein in the lumen of the endoplasmic reticulum (ER) [134-136]. In metazoans, two other proteins also act in concert with PERK in eliciting the UPR, these are inositol-requiring protein 1 (IRE1) and activating transcription factor 6 (ATF6). IRE1 activation results in auto-phosphorylation and the ability to cleave a short intron sequence from X-box binding protein-1 (XBP1) mRNA. The remaining mRNA is ligated by an unknown agent, changing its translated protein form from a repressor to an activator of transcription of UPR genes. There are also other additional mechanisms proposed for ways in which IRE1 may relieve cellular stress via interactions with tumour necrosis factor receptor (TNFR)-associated factor-2 (TRAF2) [137]. ATF6 meanwhile is transported from the ER to the Golgi apparatus. Here ATF6 is cleaved by two proteases, in the luminal domain and the intra-membrane domain, releasing the cytoplasmic domain to be

transported to the nucleus to activate UPR genes with the ER stress response element (ERSE) [138-141].

PERK, IRE1, and ATF6 all appear to be 'activated' and controlled by similar mechanisms via a highly abundant ER chaperone protein, binding immunoglobulin protein (BiP – syn: GRP78, HSPA5), a member of the HSP70 family. Under normal conditions, this 'master control protein' binds the luminal domains of PERK, IRE1 and ATF6 and prevents them from homodimerisation or transport to the Golgi. However under high levels of unfolded protein in the ER, BiP dissociates and allows PERK, IRE1 and ATF6 to undergo the relevant transformations that cause them to initiate the UPR. Several components of HCV have been shown to initiate the UPR including NS4B [142], envelope [143], and core [144] proteins and furthermore, suppression of either PERK, ATF6 or IRE1can lead to suppression of viral replication [145] indicating that the UPR is a positive factor in HCV infection. Whilst induction of the UPR, which normally leads to reduced protein synthesis, initially appears at odds with the requirement for HCV to replicate, it has been shown that HCV acts to modulate certain aspects of the UPR to ensure that viral proteins are still rapidly produced. HCV E2 and NS5A proteins have been shown to bind to and repress the kinase activity of PKR [146, 147], and the activity of the spliced isoform of XBP1 is also repressed through as yet unknown mechanisms [148] preventing downstream mechanisms that promote protein degradation. These examples of simultaneous induction and repression of highly connected pathways are

fascinating examples of the delicate control HCV exerts over cellular processes.

Since HCV is ultimately lethal to cells, much of the work to elucidate the biology of HCV infections has been done using hepatocellular carcinoma cell lines that express a sub-genomic replicon of HCV. This replicon is lacking in key components of HCV (usually the envelope protein) that prevents assembly of the complete virus particle, but still allows viral replication to occur without causing the cells to die. This also allows for much greater safety than working with viable virus particles. The proteomic studies in this chapter use Bart 79I, a highly efficient HCV replicon that has had the core, E1, E2, P7 and NS2 genes deleted [149] (Figure 4.1) and whilst the absence of these viral proteins means the model is not perfect, replicons currently represent one of the best model systems currently available for investigating new drugs to combat HCV infection.

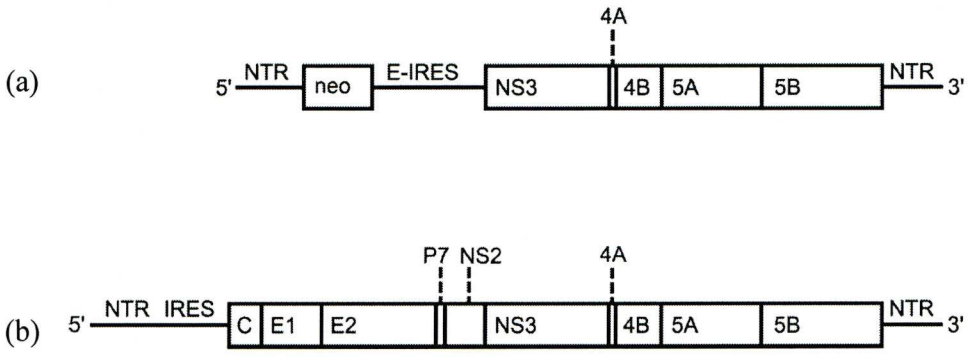


Figure 4.1. Schematic of (a) the Bart-79I replicon, and (b) the full length HCV genome. *Key*: HCV structural proteins: C = core, E1 = envelope 1, E2 = envelope 2. HCV Non-structural (NS) proteins: NS2, NS3, 4A = NS4A, 4B = NS4B, 5A = NS5A, 5B = NS5B. Other features: NTR = non translated region, neo = neomycin resistance gene, E-IRES = encephalomyocarditis internal ribosome entry site, IRES = internal ribosome entry site. Details adapted from Elazar *et al.* (2003) [149] and Tan *et al.* (2002) [150]

.

Aims

The aim of this chapter is to investigate the mode of action of the thiazolides against HCV. To achieve this Huh7 cells expressing a sub-genomic replicon of HCV were treated with NTZ and TIZ. By using DIGE to profile the proteomic response of these cells it was hoped that proteins and/or pathways involved in effecting the efficacy of the thiazolides might be revealed. Human genes and proteins are amongst the best characterised and annotated and we therefore used this data to apply three bioinformatic approaches to the investigation of possible groups of functional and biological relationships in the identified proteins.

Methods and Materials

Preparation of HCV replicon cells

RP7 cells (details below) were kindly grown, treated and lysed by Dr. Jeffrey Glenn and Dr Ping Liu, Stanford University. Huh7 cells were transformed with the Bart-79I sub-genomic replicon of HCV and grown in Dulbecco's modified Eagle's medium and 1x minimal essential medium nonessential amino acids (100x) (Invitrogen), to which 1 mg/ml of G418 was added. A G418-resistant colony colonies was isolated, expanded, passaged in 700 µg/ml G418, and termed RP7 [149]. RP7 cells were subsequently passaged in the above media with the inclusion of dimethyl sulphoxide and either 10 or 20 µM NTZ or TIZ during growth as outlined in Table 4.1.

Cell Lysis and Sample Preparation

RP7 Cell pellets were in lysed in lysis buffer consisting of 30 mM Tris/HCl, 8 M urea and 4 % (w/v) CHAPS (pH 8.5). Cell contaminants were then removed from the samples using the 2D-Quant Kit (GE Healthcare), a precipitation based method and resuspended in the same sample lysis buffer as above. Protein concentration was measured using the BSA-coomassie assay and the concentration of the samples was adjusted to 5 – 10 $mg.ml^{-1}$. The pH of each sample was checked, and if required, carefully adjusted to pH 8 – 9 using 50 mM sodium hydroxide.

2D-DIGE method

For full details of the 2D-DIGE method used please refer to the relevant portion of the Materials and Methods section in Chapter 3

CyDye Minimal Labelling

Labelling was performed as outlined in Chapter 3. Samples were randomly allocated to a gel and either Cy3 or Cy5 labelling (Table 4.1)

Gel No.	Cy2	Cy3	Cy5	Sypro
Gel 1	STD	RP7.1	NTZ10	-
Gel 2	STD	RP7.2	TIZ10	-
Gel 3	STD	NTZ20	RP7.3	-
Gel 4	STD	TIZ20	RP7.4	-
Gel 5	-	-	-	Prep

Table 4.1. CyDye labelling and allocation of protein samples to DIGE gels. Samples were labelled according to the table above. RP7.x denotes biological replicates of untreated HCV replicon expressing cells; NTZ10, RP7 cells passaged in 10 μM NTZ; NTZ20, RP7 cells passaged in 20 μM NTZ; TIZ10, RP7 cells passaged in 10 μM TIZ; TIZ20, RP7 cells passaged in 20 μM TIZ. The internal standard, STD was created from equal amounts of all samples. An unlabelled preparative load (Prep) for picking and identification purposes was made from all samples and run on a fifth gel.

Parameter	Setting
Database:	MSDB
Taxonomy:	*Homo sapiens*
Fixed modifications:	Carbamidomethylation (C)
Variable modifications:	Oxidation (M)
Enzyme	Trypsin
Maximum missed cleavages	1
Peptide charge	1+, 2+ and 3+
Peptide tolerance	± 2.0 Da
MS/MS tolerance	± 0.8 Da

Table 4.2. MASCOT parameters used to ascertain the identity of proteins modulated in DIGE experiments.

Identification of proteins by tandem mass spectrometry

MS operating procedures were identical to those outlined in Chapter 3. Database searches were performed using a dedicated in-house Mascot server, either individually using the standard ms/ms web browser based interface or using the Mascot Daemon to automate the search. In either case the parameters used are given in Table 4.2. An overall Mascot score of at least 50, and at least 70 % presence of b or y ion series in any peptide was required for a positive match. Further details including scores and peptide coverage are given in the appendix. Identifications from MSDB were mapped to several other identifiers including Entrez GeneIDs and Uniprot accessions for compatibility with other databases.

PantherDB categorisation

Proteins were grouped according to the direction of their modulation and categorised using the highest level of PantherDB classification system (www.pantherDB.org) [151] which is arranged in a directed acyclic graph (DAG) structure using increasingly more specific terms to describe a gene or protein. PantherDB possesses 31 biological processes and 29 molecular functions at its primary level of classification and is a useful tool for describing changes in experiments with large datasets. The results of the DIGE experiment were categorised using both biological process and molecular function ontologies.

DAVID gene functional classification and analysis

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID) [152, 153] (http://david.abcc.ncifcrf.gov/) was developed for the analysis of genomic data, where there is one to one representation of genes to results; however its ability to extract large amounts of data relating to the proteins or genes of interest and perform several types of meta-analysis such as the clustering of related terms to find functional groups of proteins, makes it useful for identifying features, functions, processes, and pathways in the protein data. DAVID uses kappa statistics to measure the degree of agreement between the terms annotated to pairs of genes. Pairs above specified thresholds are used to seed clusters which are then iteratively merged to form larger groups based on other rules such as the degree of overlap for members of the group. This process is repeated until clusters stabilise.

Entrez Gene IDs for up- and down-regulated proteins were uploaded to DAVID as a 'gene list' via the web submission form. The up- and down-regulated lists were also combined to create a further list of all differentially expressed proteins. For the purposes of clustering, duplicate results were ignored. DAVID annotates the submitted list with data from several selectable sources. Protein information was investigated using several different information sources however for the results outlined below, the following data sources were used to analyse this dataset: GO terms (biological process, molecular function and cellular component), protein domain information from InterPro, PIR and SMART, functional categories from SwissProt and UniProt and disease

annotations from OMIM. High stringency settings were used for the generation of clusters.

Visualisation of protein interactions using STRING

The web tool STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) [154] (http://string.embl.de/) was used to visualise possible connections between the proteins identified as being modulated upon RP7 cell treatment with the thiazolides. The network was visualised using both medium and high confidence interactions based on all available prediction methods (genomic context, high-throughput experiments, conserved co-expression and previous knowledge)

Results

Protein identifications and statistics

To learn more about the mode of action of the thiazolides against HCV, human liver carcinoma cells expressing a sub-genomic replicon of HCV were grown and passaged in the presence of either NTZ or TIZ at both 10 and 20 μM and the subsequent proteomic changes profiled using 2D-DIGE. Four biological replicates of each treatment condition and of untreated cells were requested, however only a single biological replicate of each treatment condition was received from our collaborators. As a result, thiazolide treated cells were grouped together to ensure analysis was of sufficient statistical power.

DeCyder analysis revealed 85 gel spots that were differentially expressed ($p \leq 0.05$) upon treatment with thiazolide. These protein spots were excised from the gel, and tryptically digested. Peptides were analysed by LC MS/MS and using Mascot, the ion masses were searched against the MSDB database with the *Homo sapiens* taxonomy filter applied. Of the 85 spots analysed, 70 spots contained proteins that matched confidently to a protein (or proteins) that matched the selection criteria outlined in the methods. The identifications corresponded to 40 unique proteins. The full list of protein hits is given in Table 4.3.

Entrez GeneID	Protein Name	t-test	Av. Ratio (Treated/ Control)
71	actin, gamma 1	0.024	-2.34
71	actin, gamma 1	0.044	-1.42
351	amyloid beta (A4) precursor protein	0.026	2.27
301	annexin A1	0.0018	3.82
51382	ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D	0.00043	-1.97
51382	ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D	0.012	-1.7
51382	ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D	0.028	1.71
810	calmodulin-like 3	0.0012	3.02
1192	chloride intracellular channel 1	0.0044	1.82
1152	creatine kinase, brain	0.027	2.64
51726	DnaJ (Hsp40) homolog, subfamily B, member 11	0.012	1.71
10961	endoplasmic reticulum protein 29	0.013	1.95
2023	enolase 1, (alpha)	0.037	-2.05
1936	eukaryotic translation elongation factor 1 delta	0.029	1.99
1975	eukaryotic translation initiation factor 4B	0.04	1.86
1975	eukaryotic translation initiation factor 4B	0.04	1.54
1975	eukaryotic translation initiation factor 4B	0.0066	1.65
1975	eukaryotic translation initiation factor 4B	0.023	1.92
1975	eukaryotic translation initiation factor 4B	0.0031	2.71
2950	glutathione S-transferase pi 1	0.015	1.89
2950	glutathione S-transferase pi 1	0.031	1.75
2937	glutathione synthetase	0.015	1.66
3329	heat shock 60kDa protein 1 (chaperonin)	0.0096	-3.07
3329	heat shock 60kDa protein 1 (chaperonin)	0.037	1.26
3329	heat shock 60kDa protein 1 (chaperonin)	0.017	2.36
3329	heat shock 60kDa protein 1 (chaperonin)	0.046	2.38
3329	heat shock 60kDa protein 1 (chaperonin)	0.0082	2.57
3303	heat shock 70kDa protein 1A	0.0084	3.55
3303	heat shock 70kDa protein 1A	0.014	3.88
3309	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	0.017	2.47
3309	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	0.015	3.36
3309	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	0.0024	3.44
3309	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	0.0021	3.54
3309	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	0.0071	3.91
3312	heat shock 70kDa protein 8	0.038	2.25
3312	heat shock 70kDa protein 8	0.0014	3.46
3312	heat shock 70kDa protein 8	0.041	3.82
3312	heat shock 70kDa protein 8	0.0069	4.31
3313	heat shock 70kDa protein 9 (mortalin)	0.028	2.11
3313	heat shock 70kDa protein 9 (mortalin)	0.006	2.17
3313	heat shock 70kDa protein 9 (mortalin)	0.046	2.21
3313	heat shock 70kDa protein 9 (mortalin)	0.03	2.73
3313	heat shock 70kDa protein 9 (mortalin)	0.018	3.12
3313	heat shock 70kDa protein 9 (mortalin)	0.01	1.89
3313	heat shock 70kDa protein 9 (mortalin)	0.025	3.24
3190	heterogeneous nuclear ribonucleoprotein K	0.038	-3.25
3190	heterogeneous nuclear ribonucleoprotein K	0.0085	-1.41
4001	lamin B1	0.0079	3.05
9782	matrin 3	0.015	2.47
10763	nestin	0.021	-1.78
4691	nucleolin	0.031	2.14
5184	peptidase D	0.0095	2.51
5478	peptidylprolyl isomerase A (cyclophilin A)	0.043	1.46
10935	peroxiredoxin 3	0.0063	1.63

5691	proteasome (prosome, macropain) subunit, beta type, 3	0.026	-2.69
5902	RAN binding protein 1	0.0083	2.00
5928	retinoblastoma binding protein 4	0.012	1.91
6238	ribosome binding protein 1 homolog 180kDa (dog)	0.037	1.86
6238	ribosome binding protein 1 homolog 180kDa (dog)	0.0066	2.21
-	serum albumin precursor	0.012	-2.22
6711	spectrin, beta, non-erythrocytic 1	0.0087	2.48
10992	splicing factor 3b, subunit 2, 145kDa	0.038	1.38
7086	transketolase	0.02	1.93
7086	transketolase	0.043	2.18
7086	transketolase	0.0057	2.07
7167	triosephosphate isomerase 1	0.013	-1.82
7167	triosephosphate isomerase 1	0.04	1.68
84790	tubulin, alpha 1c	0.024	1.55
10383	tubulin, beta 2C	0.015	-1.8

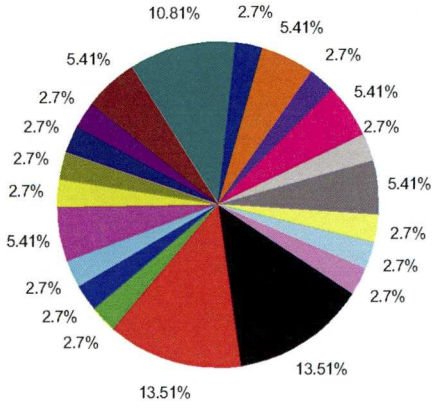
Table 4.3. Identities and expression changes of RP7 cell proteins modulated in response to thiazolide treatment. Huh7 cells expressing the Bart79I HCV, sub-genomic replicon termed RP7 were treated by the addition of either 10 μM or 20, NTZ or TIZ to their growth media (see Table 4.1) but are considered 'thiazolide treated' for statistical purposes.

PantherDB categorisation

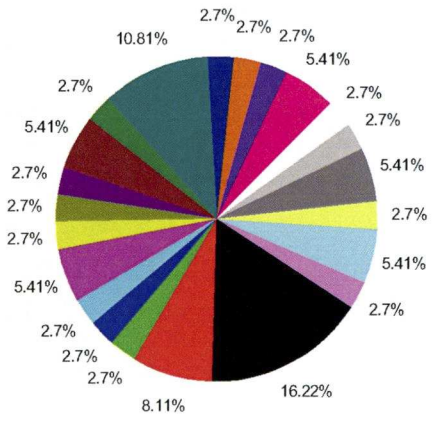
To provide a high level overview of the protein data, the PANTHER classification system was used. There are several similar classification systems designed for this purpose, such as Gene Ontologies (GO) or the MIPS Functional Catalogue (FunCat), however where PANTHER differs is the method by which classification occurs. PANTHER borrows terminology from GO terms, but rather than attempt to be comprehensive and classify each item individually, the PANTHER system develops its classifications by using a blast-based similarity scoring to create groups of proteins and to develop hidden Markov models (HMMs) for that group. Functional diversity of those high level groups is represented by HMMs developed from further subgroups of the original set. The advantage of this method is that new protein sequences can be quickly scored against the HMMs to assign the most likely appropriate functional category.

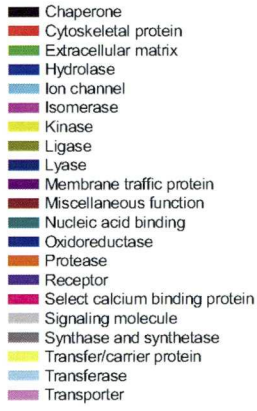
Figures 4.2 and 4.3 show the distribution of classifications for all proteins modulated by the presence of NTZ, as well as individual distributions for up-regulated and down-regulated groups. Overall, changes in the expression of 'cytoskeletal', 'chaperone' and 'nucleic acid binding' proteins represented the largest categories of molecular functional changes. The largest changes in biological processes were 'protein metabolism and modification', 'cell structure and motility', intracellular protein traffic' and 'immunity and defence'.

A



B





C

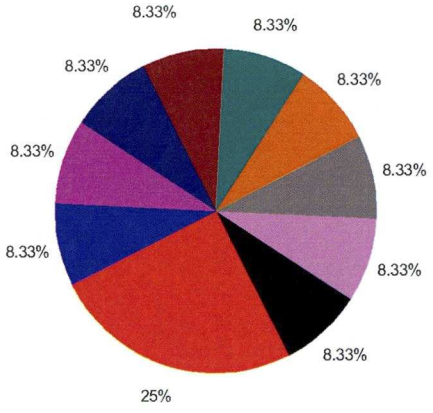
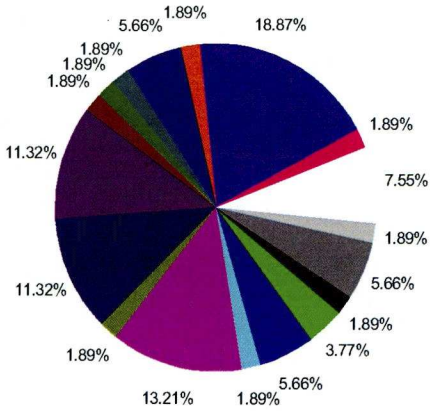
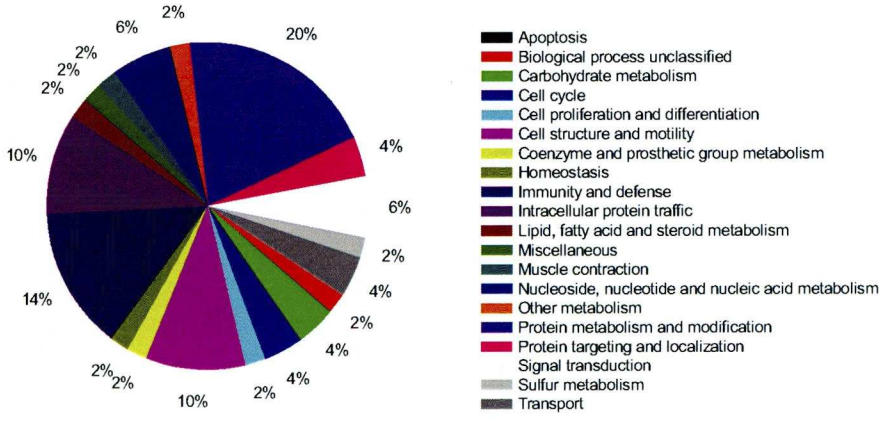


Figure 4.2. PantherDB molecular function classifications for RP7 cell proteins (A) modulated in any direction, (B) upregulated and (C) downregulated, upon treatment with the thiazolides

A



B



C

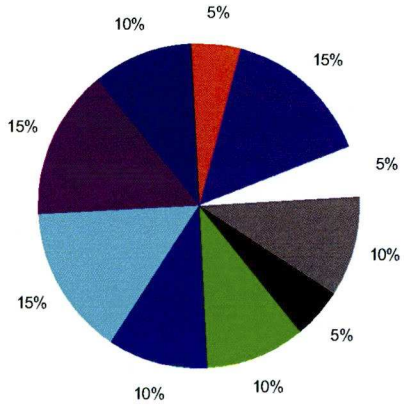


Figure 4.3. PantherDB biological process classifications for RP7 cell proteins (A) modulated in any direction, (B) upregulated and (C) downregulated, upon treatment with the thiazolides

DAVID gene functional classification and analysis

Whereas PANTHER was used to provide a simplified top-down overview of the RP7 proteins modulated in response to thaizolide treatment, the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID) [152, 153], approaches the problem from an alternative perspective. DAVID integrates all of the information from several large databases (e.g. GO, KEGG, InterPro, Entrez etc.) and uses fuzzy clustering algorithms to cluster terms associated with a set of genes based on their relatedness. This allows the user to see what groups of related functions, domains or pathways (and others) may be present in the data that may not be visible by observing protein names alone. Up-regulated, down-regulated and all modulated proteins were annotated and functionally clustered and approximations describing the most prominent resulting clusters are shown in tables 4.4, 4.5 and 4.6. The full results of the clustering analyses are included in the appendix.

Cluster Summaries for All Proteins	Max. Number in cluster
Protein chaperone	5
Anti-apoptosis	5
Cytoskeletal	11
Cell death and development	9
Cellular localisation and transport	7
Nucleotide binding (ribonucleotides)	10
Nucleoside metabolism	5
RNA binding	3
Establishment of localisation	12
Protein localisation	4
Mitochondrial transit peptide	3
Regulation of gene expression	6
Regulation of transcription	3

Table 4.4. Summary of functional clusters resulting from analysis of all proteins identified as modulated by treatment of RP7 cells with NTZ or TIZ using DAVID. Although statistical significance cannot currently be reliably counted, data is presented with higher 'enrichment scores' (i.e. enriched term with respect to that predicted by the genome, rather than proteome) at the top of the list.

Cluster Summaries for Upregulated Proteins	Max. Number in cluster
Protein chaperone	5
Anti-apoptosis	6
Protein metabolism	15
Apoptosis	7
Cellular localisation and transport	6
Nucleotide binding (ribonucleotides)	8
Establishment of localisation	10
RNA binding	3
Nucleoside metabolism	4
Protein localisation	4
Mitochondrial transit peptide	3
Ion binding	10
Regulation of gene expression	5
Regulation of transcription	3

Table 4.5. Summary of functional RP7 protein clusters up-regulated after treatment with NTZ or TIZ identified by DAVID. Although statistical significance cannot currently be reliably estimated, data is presented with higher 'enrichment scores' (i.e. enriched term with respect to that predicted by the genome, rather than proteome) at the top of the list.

Cluster Summaries for Downregulated Proteins	Max. Number in cluster
Organelle lumen	3
Structural molecule	3
Nucleotide binding	3

Table 4.6. . Summary of functional clusters resulting from the analysis of proteins that were down-regulated proteins after treatment of RP7 cells with NTZ or TIZ identified by DAVID. Although statistical significance cannot currently be reliably counted, data is presented with higher 'enrichment scores' (i.e. enriched term with respect to that predicted by the genome, rather than proteome) at the top of the list.

Visualisation of protein interactions using STRING

STRING (http://string.embl.de/) is a multispecies tool for examining known and predicted protein interactions using information derived from a number of sources varying in reliability from text-mining to expertly curated databases. The interactions returned to the user may not only be physical interactions but also associations implied by other methods such as their genomic context, or conserved co-expression. STRING is interactive and allows filtering of certain types of interactions and the various levels of interactions can be investigated in further detail and can also assign a combined probability of the interaction being real based on the number and confidences of the individual predictions. For the 41 unique identifications made, 38 could be assigned a Uniprot ID compatible with STRING. The resulting interaction network from these proteins is shown in Figure 4.4. This network consists of nodes (proteins) and edges (the interaction) and is shown in "confidence view" (Figure 4.4, top panel), where the intensity and thickness of the edges represents the combined confidence in the interaction (thicker and darker = more confident) and "interaction view" (Figure 4.4, bottom panel), with individual predictions colour coded according to their type.

The predicted STRING network is highly interconnected; indicating that the RP7 cell response detected by DIGE may well be a coordinated one rather than several disparate responses, however many of the interactions are predicted by text mining which can be unreliable. Filtration of the interaction network to only interactions with high combined confidence score (> 0.7) resulted in the network in Figure 4.5. This network appears to centre around several heat

shock and stress response proteins. GRP78 (HSPA5 in Figure 4.5) appears particularly well connected to the network. Using k-means clustering to separate the network results in three groups of proteins belonging to areas which can be approximately summarised as; ER chaperones, glycolytic enzymes and ribonucleotide binding proteins.

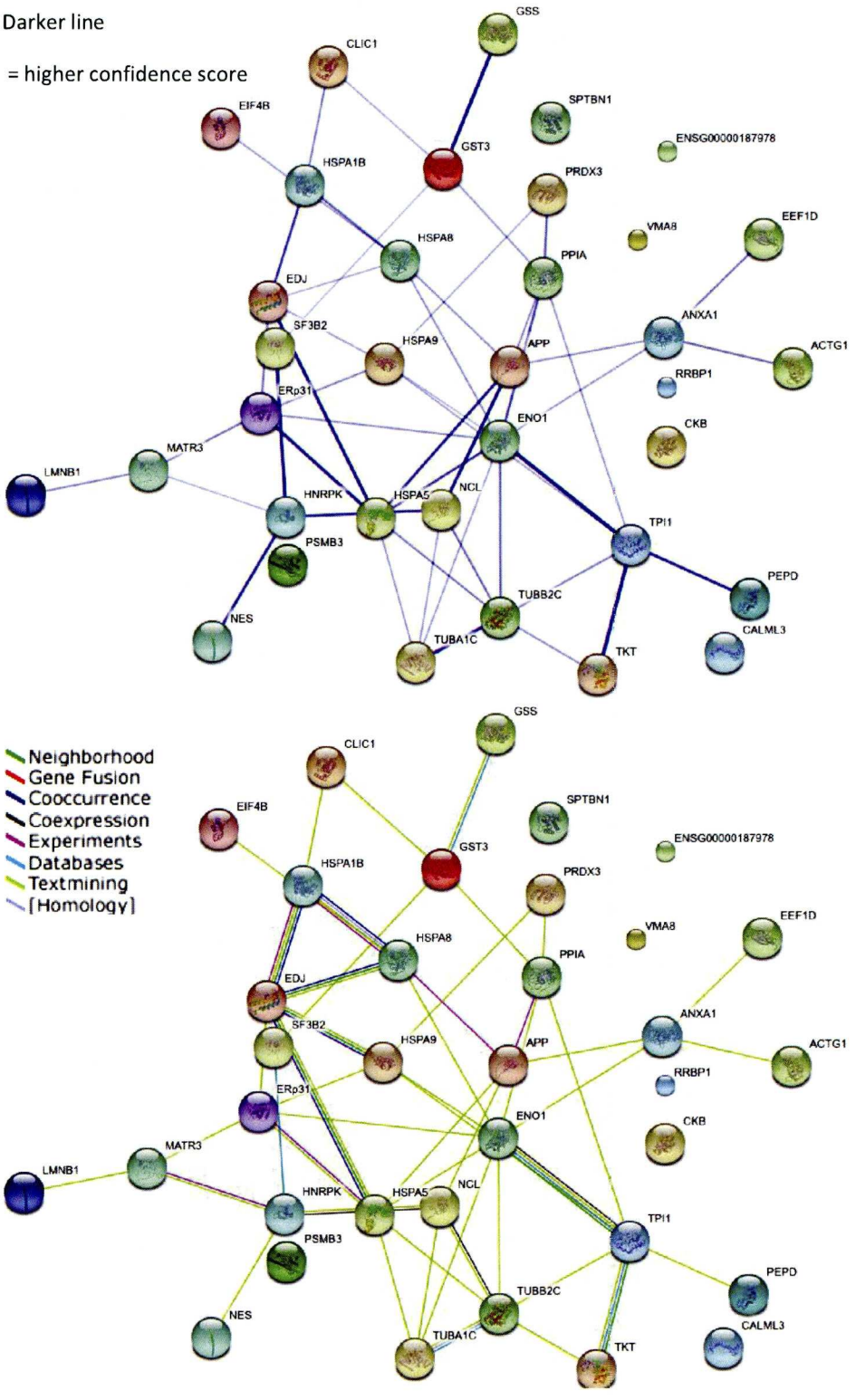


Figure 4.4. STRING protein interaction network (≥medium confidence) for RP7 cell proteins modulated in response to thiazolide treatment. Network is shown in both confidence view (top) and interaction type view (bottom).

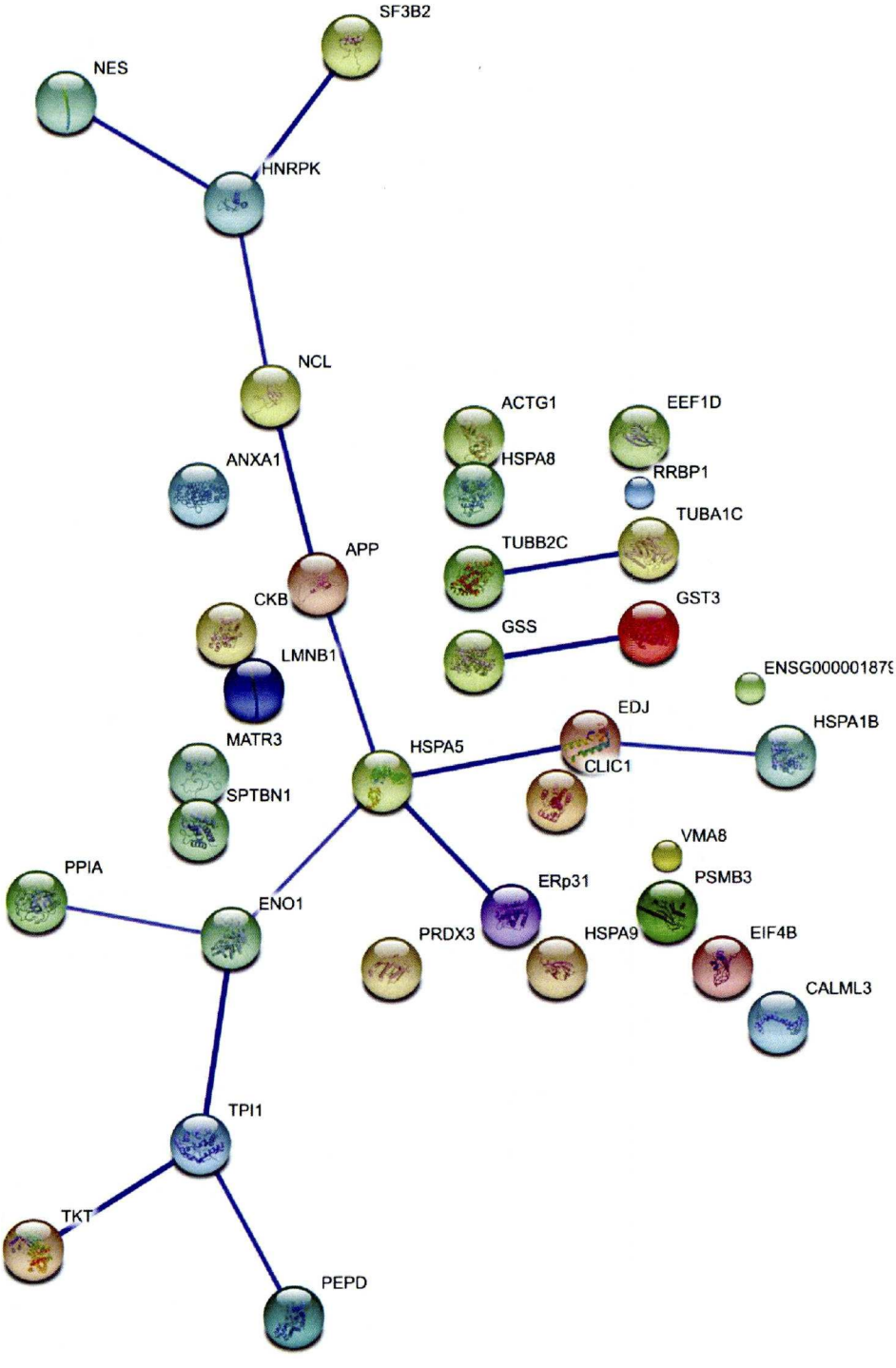


Figure 4.5. STRING protein interaction network (≥high confidence) for RP7 cell proteins modulated in response to thiazolide treatment. Network is shown in both confidence view with darker, thicker lines corresponding to higher confidence scores.

Discussion

In earlier chapters, proteomic and transcriptomic techniques were used to investigate resistance to the thiazolides in *Giardia lamblia* parasites. Some interesting avenues for further investigation were revealed, however describing the underlying biology of the system proved challenging due to the relative lack of characterisation of the *G. lamblia* genome and hence the lack of a means to integrate the transcriptomic and proteomic data in a meaningful way. Using a Huh7 cell line that had been transformed with a sub-genomic replicon of HCV (RP7cells) with clinical relevance to HCV infection, proteomic changes in response to treatment with the thiazolides were investigated. Human cell lines are well characterised and as such the annotation of the protein data should be highly beneficial in interpreting the results. Furthermore there is already some research into the proteomic response of the Huh7 cell line to the HCV replicon. In addition to looking at the individual results, three differing bioinformatics approaches were used to help describe and functionally categorise the response to the thiazolides and predict their interactions.

When RP7 cells were treated with the thiazolides and the protein separated by 2-DE the abundance of 70 identifiable protein spots corresponding to 40 unique proteins was significantly altered. After identification of these proteins, some trends were immediately apparent from the data with 24 of the 70 identifications being heat shock proteins (HSPs), and with only a single exception, all HSP spots were upregulated. This result is reflected in both the PantherDB categorisation where "chaperone" proteins are the largest

molecular function category for up-regulated proteins and joint largest in the "all regulated" protein group. Terms relating to chaperones and unfolded protein binding were also present in the DAVID cluster assignments and although applying the DAVID cluster enrichment score (designed for array type experiments) to a DIGE dataset violates some assumptions used in calculating this score, this functional cluster appeared to be the most enriched relative to the rest of the genome as the background set. Amongst the up-regulated HSPs (and some of the biggest fold changes overall) is BiP. The importance of BiP and the UPR to HCV infection has already been discussed in this chapter's introduction, and its up-regulation could provide an explanation for the up-regulation of the other HSPs. During either HCV or HCV replicon expression ATF6 is activated, cleaved and is transported to the nucleus via the Golgi to transcribe genes with ER stress response elements (ERSE) in their promoters [155]; many of these genes have chaperone functions. XBP1 is also amongst the transcriptional targets of targets of ATF6 and the up-regulation of DnaJ (Hsp40) homolog, subfamily B, member 11 (syn: HEDJ), which appears to be a transcriptional target of XBP1 [156], provides further evidence for this process. One study has also shown that BiP is transcriptionally up-regulated during HCV replicon expression, although the total BiP protein is reduced [155]. The seemingly large up-regulation of BiP upon thiazolide treatment of the RP7 cells suggests that the HCV replicon might be suppressing the translation of BiP and that the thiazolides are somehow interfering with this suppression. In fact, several of the detected changes are synonymous with the reversal of the response of Huh7 cells to

HCV or an HCV replicon. This hypothesis is somewhat supported by comparing the types of changes observed in Huh7 cells expressing an HCV replicon compared to untransformed cells [157] to the types of changes seen in the DIGE dataset. Using panther classifications, there are some similarities between proteins modulated upon introduction of the HCV replicon [157] upon treatment with the thiazolides, for example chaperones and nucleic acid binding proteins and processes such as protein metabolism and modification, however the comparison is limited by the differences in cell types between the studies. There are however several differences between these data particularly the prominence of cytoskeletal and structural proteins in this study. Both up and down-regulated cytoskeletal proteins were identified and several of these are types of actins and tubulins. Actins and tubulins have been shown to be essential to the membrane bound replication complexes (RC) of HCV [158, 159]. Furthermore we detected modulations in both peptidylprolyl isomerase A (cyclophilin A) [160] and creatine kinase [161] which have critical roles in HCV replication complexes. These results suggest that the thiazolides might act, in part, by disrupting the HCV RC. Interestingly, transketolase (TKT) which was detected as being up-regulated uses thiamine pyrophosphate (TPP) as a cofactor. As discussed in Chapter 3, PFOR is inhibited via the hijack of activated TPP by NTZ. PFOR is important to anaerobic glycolysis whilst TKT is essential to the pentose phosphate pathway that feeds into glycolysis. Triose phosphate isomerase (TPI1) and enolase (ENO1) which were both modulated upon treatment with the thiazolides are both also key components of the glycolytic pathway. Thus a mode of action that involves glycolysis should not

be ruled out; however there does not appear to be any strong direct link between glycolysis and HCV pathogenesis.

STRING was used to investigate interactions between the identified proteins. Uploading the protein IDs to the STRING web service returned a highly interconnected network with only 7 proteins not possessing a connection to the rest of the group. The interconnectedness of the group is also evident in the number of DAVID function clusters that overlapped. This is likely due to the fact that proteins such as the HSPs and structural proteins inherently interact with a great many proteins. Many of the STRING connections were however based only on evidence from text-mining and closer inspection of the network revealed that some of these connections may be unjustified. The STRING network was therefore filtered to only high confidence interactions. The resulting network appears to indicate that BiP (HSPA5 in figures 4.4 and 4.5) is well connected to a core group of proteins. However, separating the network by k-means seems to indicate three main areas exist, ER chaperones, glycolytic enzymes and ribonucleotide binding proteins. Although promising, analysis using STRING has not provided any significantly different or additional information that has not been alluded to with the other types of analysis. The potential of the technique was well demonstrated, by being the only one of the three bioinformatic techniques to isolate a group of proteins with glycolytic associations. However, with the most confident interactions belonging to well characterised diseases and processes the power to reveal novel interactions is limited. Furthermore, the parts of the resulting networks that are not supported by experimentally proven interactions may be error-

prone due to the inherent difficulties in using text-mining to predict interactions. With increased reliability of prediction algorithms, and experimental discovery of other interactions, this type of analysis may well become much more powerful and useful in reconstructing cellular processes in a systems biology manner.

The currently held hypothesis that thiazolides activate PKR through the inhibition of HCV dephosphorylation of the ER stress response enzyme can neither be fully confirmed or refuted using the data from this study but the modulation of several proteins heavily involved or related to ER stress pathways appears to lend weight to this hypothesis. The modulation of proteins required for the functioning of replication complexes of HCV could indicate another area that may be a target of the thiazolides. In Chapter 3, attempts are made to separate potential non-specific stress responses from other changes that might have relevance to the thiazolide's primary mode of action. Since it is known that HCV interacts with and controls several aspects of stress response it seems likely that attempting to distinguish between different levels of stress response would be a futile exercise.

Proteomics, transcriptomics and several bioinformatic approaches have been applied to investigate the modes of action of the thiazolides. These differential approaches can provide a great deal of information about the way some organisms respond to the thiazolides, but this information alone is unlikely to reveal the targets of the thiazolides or to prove interactions. In Chapter 4, drug targets are sought by affinity chromatography, using an untransformed (non-replicon) Huh7 cell line.

Chapter 5

Identification of Proteins with Binding Interactions to the Thiazolides

Introduction

Like many type of chromatography, affinity chromatography exploits the differences in the inherent attractive forces between interacting molecules relative to other molecules in a solution. Specifically, the user selects a molecule and immobilises it to a stationary support as 'bait' and in theory, by doing this, it should be possible to purify only those molecules that have specific affinity for the bait. Where the identification of directly interacting targets is desired, it is important to attempt to create conditions that minimise spurious interactions whilst maximising the affinity between the immobilised molecule and its true targets. In some instances, it is the ability of affinity chromatography to be performed under non-denaturing and/or native conditions that allows the 'interfering' interactions to be exploited to discover the secondary, tertiary, quaternary etc. interactions, and hence derive information regarding complexes that might be present *in vivo*. To this end there are several techniques that have sought to enhance the basic affinity technique to include very strong and/or very specific molecular interactions. Two notable and successful examples of this are tandem affinity purification (TAP tagging) which uses a custom n-terminal sequence "tagged" onto the protein of interest [162] and the use of biotin-streptavidin interactions [163, 164] where an avidin binding biotin label is conjugated to the molecule of interest. The main advantages of these methods are their very low level of contamination (especially for TAP tagging), non-

denaturing conditions and the ability to purify intact complexes. Since we neither have a known protein target to tag, nor any method for biotinylation of TIZ or NTZ, TAP tagging and biotinylation are not currently possible. Given sufficient time, it would be possible to derivatise the thiazolides to enable biotinylation and other tagging protocols. However, this was not possible in the timeframe of this project. An investigation of commercially available products that might be useful to immobilise the thiazolides revealed that the existing hydroxyl group of TIZ could be covalently attached using an epoxy activated agarose/sepharose.

There have already been several published investigations into the mode of action of the thiazolides in parasites and anaerobic bacteria including a study by Muller *et al.* [165] who used affinity chromatography to identify a novel *Giardia lamblia* nitroreductase, that can bind to and can be inhibited by NTZ. There have, however, been relatively few studies that have studied the effects of the thiazolides in the host, with one notable exception; Muller *et al.* [166] have shown that in human colon cancer cells glutathione-s-transferase pi (GSTP1) binds to and is inhibited by certain thiazolides [166]. Clinical trials are underway to test thiazolides for the treatment of HCV. It has been hypothesised that anti-HCV properties are elicited through the host cells. Since HCV affects the liver, the hepatocellular carcinoma cell line Huh7 was chosen for the investigation into potential host targets of the thiazolides. Moreover, the replicon transformed cells (RP7) used for the DIGE experiments in Chapter 4 were derived from this cell line.

Aims

The aim of this chapter was to discover which (if any) proteins are targets of the thiazolides in host cells. To acheive these aims, tizoxanide (TIZ) and two derivatives were conjugated to an epoxy activated agarose resin (EAA) and used in affinity chromatography experiments against Huh 7 cell lysates. Stearic effects that might influence the binding of putative targets to thiazolides were addressed by repeating experiments using two other thiazolides that have the same structural core as TIZ and RM4832, but benefit from the addition of an hydoxyl group on the benzene ring. It was hoped that using the alternative compounds might also give furhter insight into the way different functional groups affect the binding putative targets.

Methods and Materials

Cells and reagents

Huh7 cells grown in Dulbecco's modified Eagle's medium and 1x minimal essential medium nonessential amino acids (100x) (Invitrogen) were kindly donated by Dr. Jeffrey S. Glenn and Dr. Ping Liu (Stanford University). Frozen cells were shipped to Liverpool and lysed in approximately 1 ml of lysis buffer containing 50 mM Tris/HCl pH 8.0, 1 % (v/v) Triton-X-100, 0.1 $mg.ml^{-1}$ DNase and 2x Complete Mini protease inhibitor cocktail (Roche Diagnostics, UK) per 10^8 cells. All other reagents were purchased from VWR International (Leicestershire, UK) unless specified.

Synthesis of Thiazolides

TIZ was supplied by Romark Laboratories, L.C. (Tampa, Florida, USA). Other thiazolides were synthesized by the Chemical Synthesis Group, Department of Chemistry, University of Liverpool. A typical synthesis of RM4832 was performed as follows: To a stirred solution of acetylsalicylic acid (1.44 g, 8 mmol) in 80 mL of dry Et_2O at 0 °C was added pyridine (0.77 mL, 9.6 mmol) followed by dropwise addition of thionyl chloride (0.7 ml, 9.6 mmol). Stirring was continued at 0° C for 4 hrs, then the white precipitate was filtered off and the filtrate was concentrated under vacuum to give the acid chloride which was used without further purification. 2-Amino-5-bromothiazole, HBr salt (2.08 g, 8 mmol) followed by a solution of the above acid chloride in EtOAc (15 mL) were sequentially added to a biphasic mixture of $NaHCO_3$ (2.68 g, 32 mmol) in H_2O (25 mL) and EtOAc (25 mL) with vigorous stirring at 20°C and the

reaction was continued for 12 h. The layers were separated and the aqueous layer was extracted once with EtOAc (20 mL). The combined organic extracts were washed with 0.5 N HCl (2x 10 mL), followed by brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated under vacuum to give a pale yellow solid which was triturated with Et_2O (2x 10 mL) and filtered. The resulting solid was dried under vacuum to give the O-acetate of RM4832 (1.76g, 63%); a portion (116 mg, 0.36 mmol) was heated in conc. HCl (1mL) at 50° C for 24 hrs. The reaction mixture was cooled to ambient temperature, filtered and the solid washed with H_2O (dist.) until the washings were at neutral pH. The solid was dried under vacuum to give RM4832 (75 mg, 66%), mp 196-197°C. Found: C, 40.1; H, 2.35; N, 9.35; m/z, 298.9489; $C_{10}H_7BrN_2O_2S$ requires C40.15; H, 2.3; N, 9.3%; $C_{10}H_8BrN_2O_2S$ (MH^+ for ^{79}Br) requires m/z, 298.9490; ^1H NMR [400MHz, $(CD_3)_2SO$] 7.00 (1 H, t, J = 7.9 Hz, ArH), 7.05 (1 H, d, J = 8.2 Hz, ArH), 7.49 (1 H, t, J = 8.2 Hz, ArH), 7.65 (1 H, s, thiazole 4′-H) and 7.96 (1 H, d, J = 7.9 Hz, ArH); ^{13}C NMR [100 MHz, $(CD_3)_2SO$] 102.1, 116.3, 117.1, 119.7, 130.3, 134.5, 138.4, 157.4, 158.2 and 164.6; MS (CI) m/z 299 and 301 (M^+ for ^{79}Br, ^{81}Br respectively).

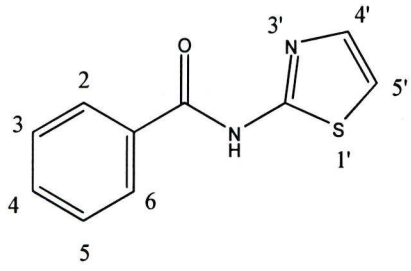
Affinity chromatography

Preparation of Affinity Chromatography Resin

TIZ, DHTIZ and DHRM4832 (see Table 5.1) were covalently attached to an epoxy activated sepharose (GE Healthcare, UK) with a 1,4-bis(2,3-epoxypropoxy-)butane (12 atom) spacer arm via either the 2- or 4- hydroxyl

residues as indicated in Table 5.1, according to the manufacturer's instructions. The procedure was as follows.

Between each of the following steps, each solution was removed by gentle centrifugation at 1000 x g and careful decanting of the resulting supernatant. Epoxy activated sepharose was washed with four rounds of distilled water to remove the storage media and to re-swell the resin, before being incubated in a volume of coupling buffer (100 mM $NaCO_3$, pH 9.5, 5 % (v/v) DMSO) equal to the volume of the swollen resin containing 200 µM of thiazolide per ml of buffer, with gentle inversion at 37 °C overnight (minimum 16 hours). The resin was then washed several times with 10 column volumes (CV) of coupling buffer to remove residual thiazolide before the medium was blocked with 1 M ethanolamine (pH 8.0) for at least 4 hours at 45°C. The blocked resin was then extensively washed with alternating rounds of high (pH 9.5) and low (pH 3) pH buffer and subsequently equilibrated with binding buffer (100 mM Tris/HCl pH 8.0, 1 % (v/v) DMSO). Control, 'drug-free' sepharose resin was obtained by following the same procedure, omitting the addition of a thiazolide.

Core Structure	Name	Chemical groups
	Nitazoxanide	**2** –O($COCH_3$) / **5'** -NO_2
	Tizoxanide	**2** –OH* / **5'** -NO_2
	DHTIZ	**2** –OH / **4** –OH* / **5'** -NO_2
	RM4832	**2** –OH / **5'** -Br
	DHRM4832	**2** –OH / **4** –OH* / **5'** –Br

*Hydroxyl groups used for conjugation with epoxy activated agarose resin in affinity chromatography experiments.

Table 5.1. Names and structures of thiazolides used or referred to in this chapter

Column Packing

Although not directly used for any particular method, the GE Healthcare Guide *"Affinity Chromatography: Principles and Methods"* [167] was used for guidance. Resin was suspended in a slurry containing approximately 10 column volumes (CV) of equilibration buffer (i.e. 1 ml resin = 10 ml buffer) and packed under gravity in a 1 cm dia. × 10 cm glass Econo-column (Bio-rad laboratories Ltd., Hemel Hempstead, UK)

Chromatography procedures

pH shift elution

The column was packed and equilibrated with 10-20 CVs binding buffer (100 mM Tris/HCl, pH 8.0). This was followed by the addition of 1 mL of Huh 7.5 cell lysate (approx 1 x 10^8 cells) clarified of particulate matter by centrifugation for 5 minutes at ~10'000 x g at 4°C before stopping the flow of the column for approximately 10 minutes to allow potentially slower binding interactions to take place. The column was then thoroughly washed with at least 10 CV of binding buffer and then 5 CV of elution buffer (100 mM glycine, pH 2.5). The column was then re-equilibrated with binding buffer before regeneration using three alternating rounds of low and high pH buffers as before. One mL fractions were continuously collected from just before addition of cell lysate to the column until at least one round of column washing was completed. Protein contained in fractions was precipitated with 10 % (w/v) trichloroacetic (TCA) acid at -20 °C overnight. Precipitated protein was recovered at ~16'000 g for 25 min at 4°C. Protein pellets were

resuspended in 30 mM Tris, 1 % (w/v) CHAPS, 6 M urea before assaying for protein concentration using Bio-Rad's BSA Coomassie assay to determine correct amounts for loading onto SDS-PAGE gels.

Competitive elution

The column was packed and equilibrated with binding buffer containing DMSO to ensure thiazolide solubility (100 mM Tris/HCl, pH 8.0, 5 % (v/v) DMSO) (10-20 CV). This was followed by the addition of 1 mL of Huh 7.5 cell lysate before stopping the flow of the column for approximately 10 minutes to allow potentially slower binding interactions to take place. The column was then thoroughly washed with 20 CV of binding buffer. Five CV of competitive elution buffer (100 mM Tris/HCl pH 8.0, 25 % (v/v) DMSO, 1 mM TIZ) was added to the column and 1 ml fractions collected in 1.5 ml microcentrifuge tubes from just before addition of cell lysate to the column until the start of column washing steps commenced. This was followed by a re-equilibration with a low pH elution buffer (100 mM glycine pH 3.0, 1 % (v/v) DMSO) before regenerating the column using alternating rounds of high and low pH buffer. Fractions were collected and protein was concentrated by TCA precipitation for separation and further analysis on 12 % (w/v) 1D-SDS PAGE gels.

Pre-incubation with TIZ

Prior to chromatography, approximately 1 ml of cell lysate (0.25×10^8 cells) was treated with 1 mM TIZ for 4 hours at room temperature (≈21 °C), before

chromatography was performed using the competitive elution method described above.

Protein recovery and concentration

RP trap column

In initial experiments unsuccessful attempts were made to concentrate the relatively dilute eluate using a small reversed phase column. Eluted fractions were removed of particulate matter using a 45 µm pore syringe filter (Millipore). A C4 reversed phase micro trap column (Michrom Bioresources, Inc., California) was equilibrated with 2 % (v/v) acetonitrile (ACN), 97.9 % (v/v) H_2O, 0.1 % (v/v) trifluoroacetic acid (TFA). Fractions were brought to the same ACN/TFA composition before being injected over the column. A second buffer of 70 % (v/v) ACN, 29.9 % (v/v) H_2O, 0.1% (v/v) TFA was used to elute proteins from the column.

TCA precipitation

To concentrate protein from the dilute eluate for visualization on SDS-PAGE gels, 100 % (w/v) TCA was prepared and to each 1 ml fraction, 111 µl was added to give a final concentration of 10 %. Tubes were mixed by inversion and centrifuged for 30 minutes at 16,000 x *g* at 4 °C to sediment the protein. The supernatant was carefully removed and discarded whilst the protein pellet was resuspended in 15 µl of 30 mM Tris/HCl, 4 % CHAPS, 8 M urea and the pH was carefully adjusted to pH 8.0 with the addition of 2 M.

1D SDS-PAGE

Samples in a modified Laemmli sample buffer [168] (120 mM Tris/HCl, 20 % (v/v) glycerol, 4 % (w/v) SDS, 200 mM DTT, trace Pyronin Y), were heated at 95 °C for 5 minutes before being loaded onto a 12 % (v/v) acrylamide gel with a 4 % (v/v) acrylamide stacker. Electrophoresis was performed at a constant 200 V for approximately 45 minutes. Benchmark Protein Ladder (Invitrogen) was used as the molecular weight marker throughout. Gels were post-stained with Sypro Ruby (Invitrogen) and visualised using the Ettan DIGE Imager (GE Healthcare) followed by staining with colloidal coomassie for manual excision of bands for MS identifications. The 'quick' Sypro Ruby staining method was used as outlined in the manufacturer's instructions. Gels were washed with gentle agitation in a fixing solution of 50 % (v/v) methanol, 7 % (v/v) acetic acid for 30 minutes at room temperature with fresh buffer added after 15 minutes. The fixing solution was subsequently replaced with Sypro Ruby stain and the gel was heated in a microwave to just below the boiling point of the solution, this heating was repeated after leaving gel to cool for 30 seconds and again after 5 minutes with gentle agitation between heating steps. This was then agitated for a further 23 minutes. The gels were then washed for 30 minutes at room temperature with a mixture of 10 % (v/v) methanol and 7 % (v/v) acetic acid.

In order to facilitate the excision of protein bands, the gels were post-stained with colloidal coomassie after brief washing with distilled water for 5 minutes before immersing the gels in a solution of 20 % (v/v) methanol, 0.08 % (w/v) coomassie brilliant blue (G250), 0.01 % (v/v) phosphoric acid, 10 % (w/v)

ammonium sulphate for 1 – 7 days until protein was sufficiently stained. The gels were then washed with distilled water to reduce background staining of the gels and stored in 1 % (v/v) acetic acid solution.

Protein identification by mass spectrometry

Tryptic digestion of proteins

Gel bands containing proteins identified as being of interest were excised and destained with repeating rounds of 25 mM ammonium bicarbonate, 50 % (v/v) acetonitrile (ACN) for 10 min at 37 °C, then 100 % (v/v) ACN for 10 min at 37 °C until completely colourless. Solvent was removed and the plug allowed to air dry for 10 min. 10 µl of Trypsin solution (25 µg Trypsin diluted with 250 µl 50 mM Acetic acid then diluted 1 in 10 with 25 mM ammonium bicarbonate) was added and incubated at 37 °C for 1 hr followed by the addition of a further 10 µl 25 mM ammonium bicarbonate and incubation at 37°C overnight. Digestion was stopped with 2 µl of 2.6 M formic acid.

Mass Spectrometry

LC-MS/MS data of tryptic peptides was acquired using a LTQ ion-trap mass spectrometer (Thermo-Electron, Hemel Hempstead, UK) coupled on-line to a Dionex Ultimate 3000 (Dionex Company, Amsterdam, The Netherlands) HPLC system equipped with a nano pepMap100 C18 reversed phase column (75 µm; 3 µm, 100 Angstroms) as described in the General Materials and Methods chapter.

Results

Conjugation of thiazolides to epoxy activated sepharose

The manufacturer's instructions state that 1 ml of swollen resin contains 25-40 µmoles of active ligand, thus the concentration of 200 µmoles of TIZ per ml of coupling solution was expected to be in excess. There is currently no simple, validated method for assaying the amount of TIZ that had bound to the agarose, however, at the pH ranges used during this procedure, TIZ, DHTIZ and DHRM4832 are bright yellow/orange in solution and since this colour remained after extensive washing with buffers containing high concentrations of DMSO (to ensure complete solubilisation of compounds), it was assumed that adequate binding to the epoxy activated agarose had occurred.

Concentration of column eluate using a reversed phase trap

The eluate from the affinity columns was too dilute to be analysed by 1D SDS-PAGE without prior concentration. By using a reversed phase protein trap column it was hoped that protein could be efficiently recovered and concentrated without resorting to protein precipitation which can result in protein that is difficult to resolubilise. Unfortunately, this method repeatedly resulted in total loss of protein as determined by visualisation with 1D SDS-PAGE (data not shown). This procedure method was abandoned in favour of a TCA precipitation method.

Elution of bound proteins by decreasing pH

The chromatography procedure resulted in a large number of fractions that were too dilute for futher analysis. To determine which fractions contained eluted proteins (if any), the elution profile of the chromatography procedure was determined. Fractions were TCA precipitated and protein concentration quantified; a representative elution profile from a low pH elution experiment can be seen in Figure 5.1. In all five experiments of this kind, the elution profiles showed similar peaks at flowthrough (fractions F1 – F4) and for elution (E2). These peak fractions were visualised using 1D-SDS PAGE Although an initial trial-run lacking proper controls appeared promising, with several bands that appeared enriched in the eluate compared to the raw lysate and flowthrough fractions, subsequent analysis with the inclusion of proper controls (Figure 5.2) consisting of ethanolamine blocked EAA without tizoxanide using different quantities of cell lysate, highlighted a particular problem with the specificity of the resin and effectively excluded several of the previous "candidates" for TIZ binding. The high level of interfering interactions in this system made it very difficult to distinguish true interactions with TIZ from interactions with the agarose resin or spacer arm. It further highlighted the requirement for a more stringent or thorough washing and a more specific method of elution. One band of approximately 24-25 kDa does however appear to be consistently eluted from TIZ-agarose but not control-agarose resins. MS analysis of total eluate indicates this band is likely to be GSTP1, a previously identified thiazolide binding protein [166].

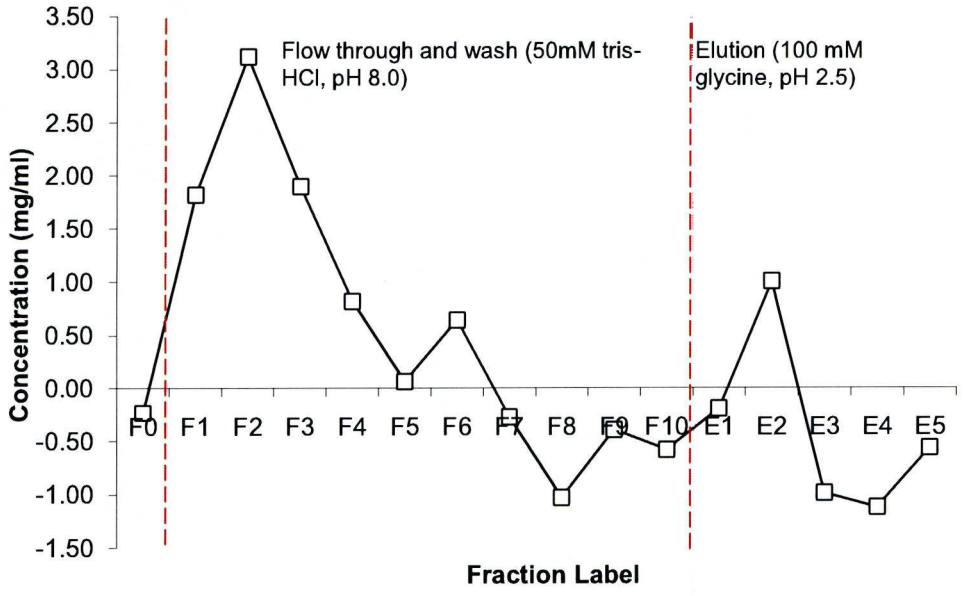


Figure 5.1. Typical elution profile of Huh7 cell extract using TIZ-agarose. Chromatography was performed in two main steps; a binding and washing step, followed by a high pH elution step. Fractions were collected and TCA precipitated before assaying them for protein concentration using Bio-Rad BSA Coomassie assay.

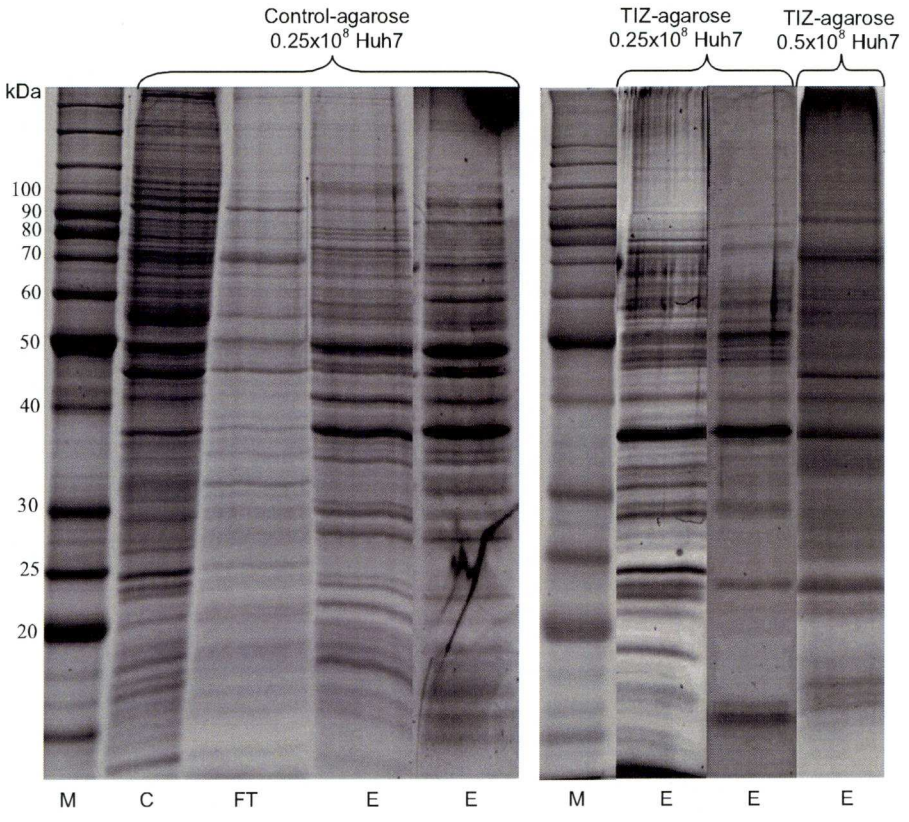


Figure 5.2. SDS PAGE analysis of Huh7 cell proteins eluted from TIZ-agarose and unbound agarose columns. Separations of Huh7 cell lysate were performed using both control-agarose and TIZ –agarose columns using glycine elution buffer (pH 3.0) to remove bound proteins. Lanes: M = Mw (kDa) marker, C = cell lysate, FT = Flowthrough after washing steps, E = protein eluted at pH 3.0.

Competetive elution with NTZ

To try and improve specificity of elution, more extensive washes of 'unbound' cell lysate were introduced to remove as many non-specific interactions from the column as possible. Increased washing appeared to have had a positive effect on the level of background binding and was maintained throughout subsequent experiments. A 1 mM TIZ wash was also used to selectively release thiazoldie bound proteins from the column, further enhancing specificity of the results.

In addition to these changes, it was decided that a more comprehensive overview of the elution might provide useful information, and so all fractions were TCA precipitated and visualised using 1D SDS-PAGE (Figure 5.3). Background subtracted lane profiles were also generated using TotalLab (NonLinear Dynamics Ltd.). One main band corresponding to a protein of ~31 kDa was consistently eluted by 1 mM TIZ elution buffer when using the TIZ agarose (Figure 5.3, lane 18, and close-up detail Figure 5.4, panels A and B). This protein band was not seen with blocked control agarose resin which lacked conjugated TIZ (Figure 5.4, panel C), indicating that the interaction was not a result of non-specific binding. This band was excised from the analysis of two separate separations and digested with trypsin before analysis by LC MS/MS. In both cases, there were several proteins identified in the gel slice, however in both cases the highest scoring protein was identified as NAD(P)H quinone oxidoreductase (NQO1) with highly significant Mascot scores and good sequence coverage (see Figure 5.5 and appedices). The

monomer of NQO1 has a mass of 30774 Da which corresponds well to the mass of the band observed on the gels.

A second band of approximately 38 kDa also appears to follow a similar elution profile over the first four elution fractions (Figure 5.4, fractions 18-22), however this behaviour was also observed in control experiments and was therefore an identification was not sought.

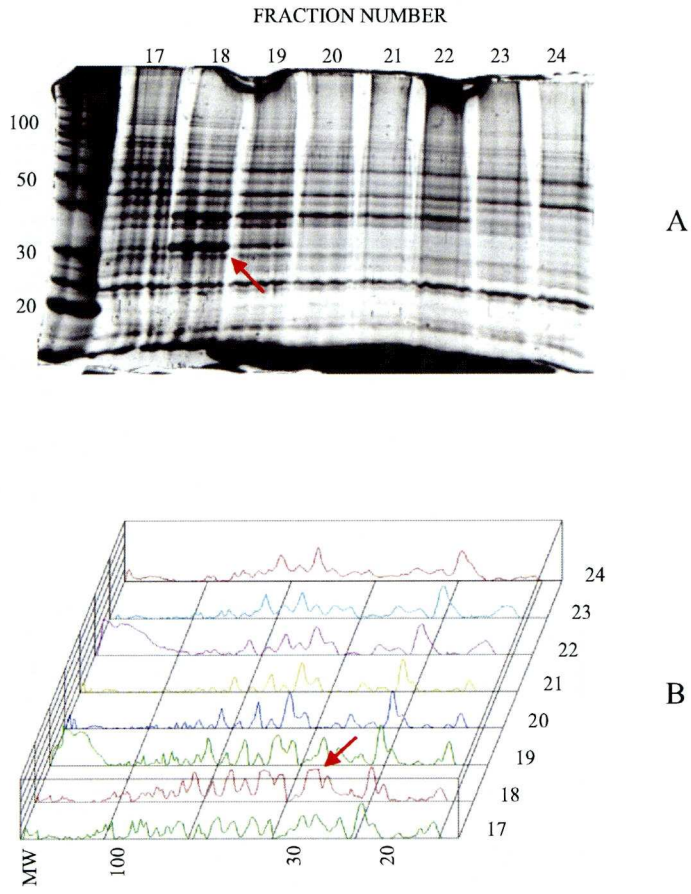


Figure 5.3. Visualisation of protein fractions eluted from TIZ-agarose column using competitive elution. Fraction 17 – flowthrough. Fractions 18 - 24 fractions eluted with 1 mM TIZ. Approximate molecular weights (MW) in kDa. Arrows indicate band excised for identification by MS. A) Original gel image B) Background subtracted and gel distortion corrected lane profiles analysed using TotalLab.

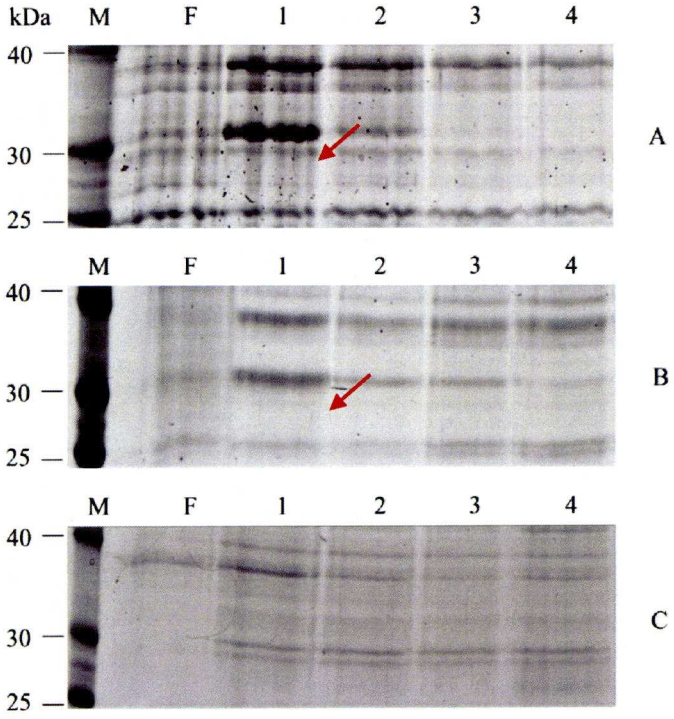


Figure 5.4. Close-up detail of SDS-PAGE visualisation of competitive elution using TIZ containing buffer. A and B) using TIZ-EAA resin and C) using ethanolamine blocked control resin EAA. The bands at approximately 31 kDa in panels A and B as indicated with arrows, were unambiguously identified by LC-MS/MS as NAD(P)H Quinone oxidoreductase 1.

Match (ID):	MASCOT Score:	Sequence coverage (%):
NQO1_HUMAN	343 / 357	58 / 48 (Combined: 60)

Sequence and peptide coverage

```
1      VGRRALIVLA HSERTSFNYA MKEAAAAALK KKGWEVVESD LYAMNFNPII
          ******* ********** **********  ********* **********

       **  ****** ********** ********** ********** **********
51     SRKDITGKLK DPANFQYPAE SVLAYKEGHL SPDIVAEQKK LEAADLVIFQ
       **       ** ********** **********  ********* **********

                     ******* ********** ********
101    FPLQWFGVPA ILKGWFERVF IGEFAYTYAA MYDKGPFRSK KAVLSITTGG
       ********** ***      ** ********** ********

151    SGSMYSLQGI HGDMNVILWP IQSGILHFCG FQVLEPQLTY SIGHTPADAR

       *********
201    IQILEGWKKR LENIWDETPL YFAPSSLFDL NFQAGFLMKK EVQDEEKNKK
       *********                                 * *******

       ********** **********
251    FGLSVGHHLG KSIPTDNQIK ARK
       ********** **********
```

Figure 5.5. Identification of NQO1 by mass spectrometry. Asterisks below and above the amino acid sequence indicate hits in the first and second identifications respectively

Alternative conjugates

The stoichiometry and stereochemistry of binding are likely to be important factors in any potential interactions with proteins, particularly as the ligand has been immobilised using a spacer arm which will undoubtedly affect any active site interactions to some extent by imposing spatial constraints. To investigate this, working with Mazhar Iqbal in the Dept of Chemistry, two compounds were synthesised that possessed alternative binding sites for conjugation to the epoxy activated agarose (Figure 5.6). Competitive elution chromatography was conducted in the manner described previously and the SDS-PAGE results of the final flowthrough fraction and first elution fractions are shown in Figure 5.6 (bottom panel). With these conjugates there do not appear to be any bands specifically eluted by TIZ, and the lack of a band that corresponds to the elution of NQO1 as in previous experiments was also noted. A similar experiment by our collaborators revealed that NQO1 also binds to another thiazolide that is structurally similar to DHRM4832 with a methyl group in place of the secondary hydroxyl group (Figure 5.7) where binding occurs in the same orientation as TIZ.

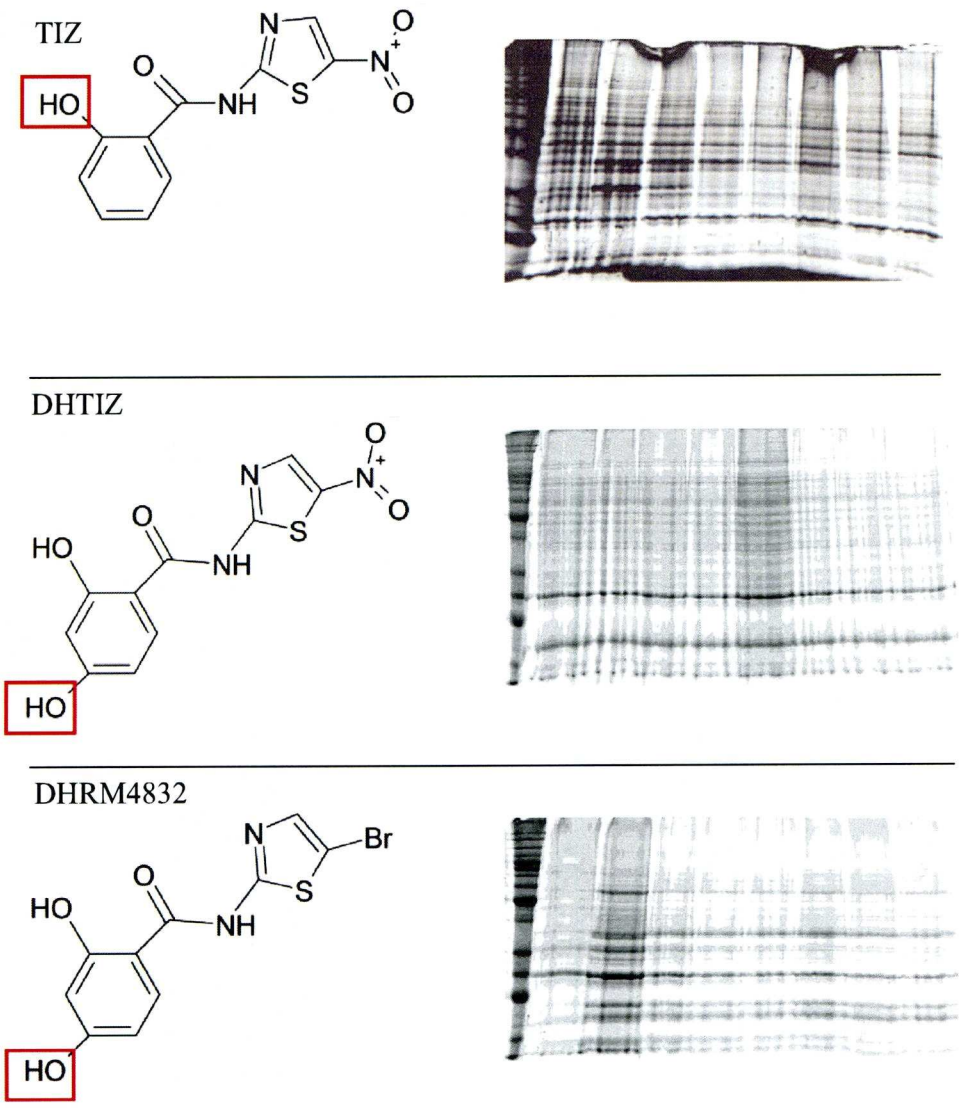


Figure 5.6. Structures of TIZ and alternative conjugates DHTIZ, DHRM4832 and SDS-PAGE analysis of affinity chromatography using these conjugates. Red boxes indicate most probable point of conjugation to epoxy activated agarose. Gel lanes (from left to right) Marker, final wash, first eluate fraction, second eluate fraction, third etc.

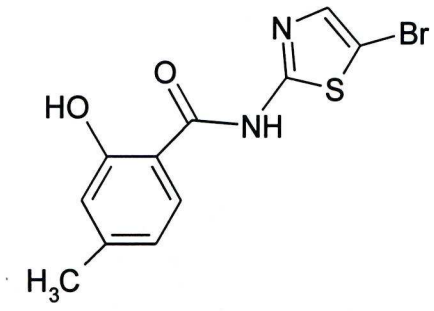


Figure 5.7. The structure of the NQO1 binding thiazolide. The above thiazolide was conjugated to an EAA resin via the hydroxyl group.

Pre-Incubation with TIZ

Although a novel putative target of TIZ had been identified (NQO1) by elution with TIZ, the identification ultimately relies on interactions with a conjugated thiazolide that is unlike that which would be encountered *in vivo*. To address this and to circumvent any issues of non-specific binding, Huh7 cell lysate was preincubated with TIZ so that binding sites could be occupied by the drug and therefore prevent binding to TIZ-agarose. This was compared to an untreated cell lysate subjected to the same procedure. Figure 5.8 shows proteins competitively eluted from TIZ-agarose for both treated and untreated cell lysates. There were no observable differences between the two conditions, even in the region corresponding to the expected location of NQO1, for which binding has already been demonstrated. Whilst there may be several factors that have prevented the observation of differential binding between the two conditions, the lack a band corresponding to NQO1 may indicate that non-specific proteolytic degradation has occurred. Despite the inclusion of a protease inhibitor cocktail in the lysis buffer, this is still possible since some commonly used protease inhibitors (e.g. PMSF) posses half lives in the order of an hour at room temperatures.

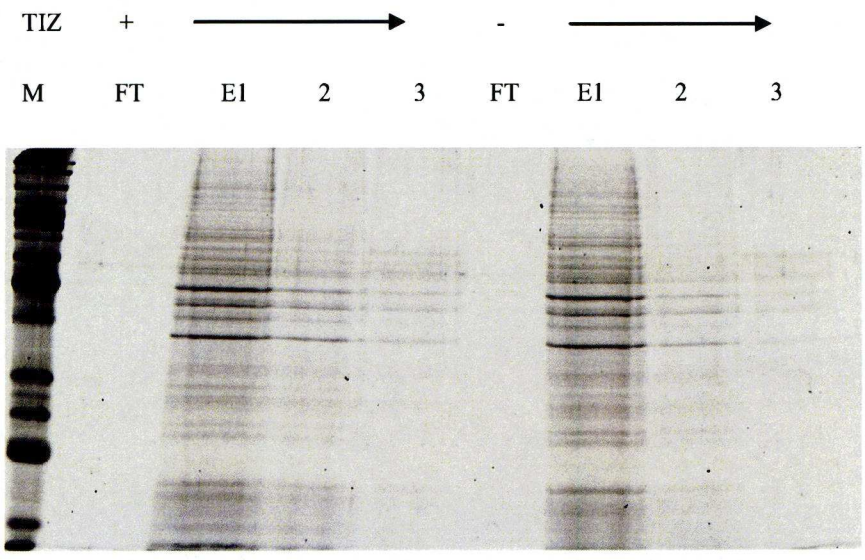


Figure 5. 8. Comparison of competitively eluted proteins bound to TIZ-agarose after with (+) and without (-) pre-treatment of Huh7 cell lysates with TIZ. Lanes: M = Mw Marker, FT = final flowthrough after washing step, E1, 2 and 3 = 1 mM TIZ eluted fractions.

Discussion

Identification of cellular targets of new drugs is an essential part of understanding their mode of action and important in improving their efficacy and safety. The application of affinity chromatography to this area of research has both advantages and disadvantages, but when working effectively provides a targeted approach to answering complex biological questions (e.g. [169]). In combination with other experimentation (e.g. transcriptomics, proteomics, or classical biochemical studies), the relevance of an interaction can be assessed.

Conceptually, the method is simple; the inherent and specific affinity between an immobilised drug and its targets are used to capture binding partners from a homogeneous solution of biological sample (cell lysate). There are however a number of areas where contamination of the results with interfering data such as non-specific binding and 'piggybacking' interactions makes interpretation of the results difficult and the analysis problematic. Non-specific binding with the chromatography bed is probably most problematic in this study however. As can be clearly seen in Figures 5.2, and 5.4, panel C, even with blocked agarose containing no bound ligand, there is a remarkably high level of protein binding to the chromatography bed, eliminating several strongly staining bands that would otherwise appear to be clear candidates for targets of TIZ. Since contamination was present in both control and TIZ-agarose resins, the background binding must be a product of methodological and technical factors, most probably interactions with the porous agarose and long spacer arm (a 12 carbon 1,4-bis(2,3-epoxypropoxy-)butane) [170, 171]. Similar problems with background binding can be seen in a 2007 study of *Giardia*

lamblia, using the same type of chromatography bed conjugated to the thiazolides [165].

Nevertheless, with improvements that enable deconvolution of interfering data from the sought after biological result, affinity chromatography has proven itself to be a valuable tool for identifying possible cellular targets of the thiazolides, reliably identifying NQO1 as a binding partner to TIZ, however if further information is to be gleaned or more research is to be conducted using other conjugates, improvements in the specificity of the separation must still undoubtedly be made. Improvements in removing the background contamination would likely be gained by moving away from a gravity flow method and using either a column packed under pressure with fully controllable linear flow rates and real-time absorbance measurements or by using a non-column method such as batch chromatography or even a hybrid method such as expanded-bed chromatography [172] that would allow complete suspension of the resin in the washing buffers. Developments should also be made in the method for concentration of the column elute for visualisation by SDS-PAGE. Initially, a reversed phase protein trap column was also investigated, this column was used within recommended parameters and manufacturer's instructions were followed precisely. However this method repeatedly resulted in the total loss of protein. The exact cause of this failure is currently unknown; however the low pH of the elution buffer and the small amounts of TFA in the concentrated protein may have rendered samples incompatible with SDS-PAGE. It is also possible that a fault in the trap holder which might have caused sample to leak to waste, or that no protein was eluted

from the TIZ-agarose column during these experiments. Since manual operation of the trap column was also proving to be time consuming, the RP trap was abandoned in favour of a more straightforward TCA precipitation with the appropriate pH adjustments. However, particularly if used in-line with mass spectrometry, the RP trap column method could still prove useful. Other methods such as lyophilisation [173], dialysis and centrifugal concentration [174-176] could also be tested for suitability for this application.

Other methods of investigation were also concieved; working with Mazhar Iqbal (Dept of Chemistry) plans to derivatise the thiazolides and enable the use of click chemistry [177] were made. This would have enabled the use of a range of covalent tags such as biotin. It was also hoped that thiazolides bearing an azido group could also be synthesised. This photoreactive bond would have enabled the thiazolides to be covalently linked to their *in vivo* targets and allowed for some interesting MS mass tagging applications using heavy and light isotopes, and allowed for experiments to determine the site of binding in any elucidated interactions.

Ultimately affinity chromatography has the power to identify possible targets of a drug, however to determine if this interaction is biologically meaningful, direct investigation of this putative interaction must be performed. In Chapter 6, the interaction between NQO1 and the thiazolides is further investigated by classical enzyme kinetic techniques.

Chapter 6

Enzyme Kinetics of NQO1 and the Thiazolides

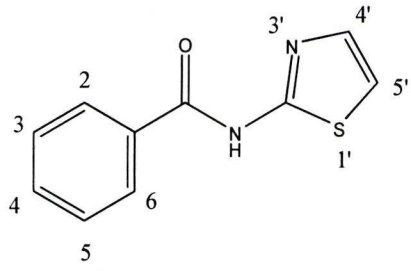
Introduction

In Chapter 5, affinity chromatography of Huh cell lysate using three different immobilized thiazolides, demonstrated that there was an interaction between the immobilised form of Tizoxanide (TIZ) and human NAD(P)H Quinone oxidoreductase 1 (NQO1). Furthermore, this interaction was not apparent with the two other thiazolide conjugates that were used for chromatography. These thiazolide derivatives were synthesized with an additional hydroxyl group on position 4 of the benzene ring moiety in order to provide an alternative orientation for conjugation with the agarose resin.

NQO1, a homodimeric FAD-binding protein, catalyses the two electron reduction of quinones to their less toxic hydroquinone form, bypassing radical formation that might otherwise damage the cell. Undergoing what has been termed a 'ping-pong-type' reaction NADH binds the active site and reduces FAD, a cofactor situated close-by. NAD^+ then leaves the active site whereupon quinone can bind and subsequently be reduced to its hydroquinone form before release. Access to the active site is carefully controlled by tyrosine-128 and residues 232-236, which form a loop, that act as a lid and restrict access to the binding after either the substrate or cofactor has departed [178]. NQO1 is an important cytoprotective enzyme, whose response is mediated by the antioxidant response element (ARE) and is also involved

more generally in xenobiotic metabolism, for example the detoxification of the phenolic metabolites of benzene [179]. NQO1 also interacts with and stabilises p53, an anti-apoptotic protein [180]. Mutations in p53 are often associated with certain types of cancers. Altered expression of NQO1 has been observed in several aberrant tissues [181], including several tumour cells, making it an interesting target for pro-drug type treatments for cancer whereby the drugs undergo what has been termed 'bioreductive alkylation' by NQO1 to become 'activated' and as a consequence causing toxicity to the cell [182, 183]. NQO1 expression is typically induced by oxidative stress, via the induction of Nrf2 and subsequent binding to genes possessing ARE in their upstream sequences [184].

To determine if the interaction between the thiazolides and NQO1 may be of physiological significance, the interaction between the thiazolides and NQO1 was further studied. The binding interactions of TIZ and five other thiazolides were investigated using in solution kinetics and *in silico* modelling approaches.

Core Structure	Name	Chemical groups
	Nitazoxanide	2 –O($COCH_3$) / 5' -NO_2
	Tizoxanide	2 –OH* / 5' -NO_2
	DHTIZ	2 –OH / 4 –OH* / 5' -NO_2
	RM4819	2 -OH / 3 -CH_3 / 5' –Br
	RM4832	2 –OH / 5' -Br
	DHRM4832	2 –OH / 4 –OH* / 5' -Br
	RM4850	2 -OH / 5 -CH_3 / 5' -Cl
	RM4857	2 -OH / 5' –H
	RM4863	2 –OH/ 5' –SO_2CH_3

*Hydroxyl groups used for conjugation with epoxy activated agarose resin in affinity chromatography experiments.

Figure 6.1. Names and structures of thiazolides used or cited in this chapter.

Aims

The experiments of Chapter 5 demonstrated that TIZ binds to NQO1 and that it is eluted under low millimolar concentrations. The aim of this chapter was to examine in greater detail the binding interactions between NQO1 and several thiazolides, including TIZ, using in solution kinetics and *in silico* modelling approaches to determine its possible physiological significance. To achieve this, an assay measuring the enzymatic activity of NQO1 was established and the effects of thiazolides on the activity were assessed. Computational modelling experiments were also undertaken to predict how interactions might be mediated.

Methods and Materials

Cells and reagents

Huh 7.5 cells, which were kindly donated by Jeffrey S. Glenn (Stanford University), were lysed in buffer containing 50 mM Tris pH 8.0, 1 % (v/v) Triton-X-100, 0.1 $mg.ml^{-1}$ DNase and 2x Complete Mini protease inhibitor cocktail (Roche Diagnostics, UK). All other reagents were purchased from VWR International (Leicestershire, UK) unless specified.

Sources of thiazolides

TIZ was supplied by Romark Laboratories, L.C. (Tampa, Florida, USA). Other thiazolides were synthesized by the Chemical Synthesis Group, Dept of Chemistry, University of Liverpool, under the supervision of Dr A. Stachulski.

Initial assay conditions for monitoring quinone reduction.

The initial conditions were based on Ernster [185]. The final reaction mixture contained 50 mM Tris/HCl pH 7.4, 5% DMSO (for drug solubility), 200 µM NADH, 65 µM menadione sodium bisulphite (MBS), 1 unit of recombinant NQO1 in a final reaction volume of 1 mL. The reaction was started by adding the reaction buffer to MBS and NQO1 and subsequent mixing by inversion. A positive reaction was shown by the conversion of NADH to NAD+ detected at 340 nm using an Ultrospec 4100 (GE Healthcare) at 21 °C. Concentrations of NADH were calculated using an extinction coefficient of 6220 $M^{-1}.cm^{-1}$

Troubleshooting of Assay Conditions

During steps taken to troubleshoot the assay, the reaction buffer, based on the work of Lind *et al.* [186], contained the following: 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 200 μM NADH and starting amounts of NQO1and MBS were varied as specified in the results. The reaction was monitored by the removal of an aliquot of reaction mixture and measurement of the conversion of NADH to NAD+ at 340 nm using an Ultrospec 4100 (GE Healthcare) at 21 °C. Concentrations of NADH were calculated using an extinction coefficient of 6220 $M^{-1}.cm^{-1}$

Monitoring of reaction via reduction of cytochrome C.

Since the previous efforts to use NADH at 340 nm to monitor the reduction of menadione bisulphite to its hydroquinone form had been unsuccessful, a new assay was adapted from Lind *et al.* [186]. The reaction was assayed in 50 mM Tris/HCl pH 7.4, 500 μM NADH, 0.08 % (v/v) Triton-X-100 (an activator of NQO1), 0.07 % (w/v) bovine serum albumin (also an activator of NQO1) (Sigma Aldrich, UK), 77 μM cytochrome C (Sigma Aldrich, UK) in a final reaction volume of 1 mL. Menadione (MEN) and NQO1 were added to the reaction as specified in the results section. The reaction was started by the addition of NQO1 (Sigma Aldrich, UK). Reaction rates at 21 °C were determined using menadiol-mediated reduction of cytochrome C, which was monitored by an ultrospec 4100 (GE Healthcare, UK) at 550 nM. An extinction coefficient (ε) of 29.5 $mM^{-1}.cm^{-1}$ was used to determine concentration of reduced cytochrome C.

Final NQO1 $K_{m'menadione}$ and IC50 determinations

The NQO1 enzyme assay was adapted from Lind *et al.* [186]. The reaction was assayed in 50 mM Tris/HCl pH 7.4, 500 μM NADH, 0.08 % (v/v) Triton-X-100, 0.07 % (w/v) bovine serum albumin (Sigma Aldrich, UK), 77 μM cytochrome C (Sigma Aldrich, UK) and menadione over a range of concentrations (0.125, 0.25, 0.5, 1, 2, 4, 8, 10 μM) in a final reaction volume of 1 mL. The reaction was started via the addition of 50 ng (0.0246 U) human recombinant NQO1 (Sigma Aldrich, UK). The menadiol mediated reduction of cytochrome C was monitored at 550 nM using an ultrospec 4100 (GE Healthcare, UK). An extinction coefficient of 29.5 $mM^{-1}.cm^{-1}$ was used for cytochrome C. K_m was calculated by non-linear regression using Graphpad's Prism software. For IC50 determinations, reaction conditions were as above but the menadione concentration was fixed at 2 μM and varying concentrations of thiazolide (0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30 μM) were used. IC50 values were calculated using nonlinear regression using PRISM (Graphpad Software).

Computational modelling

Computational modelling was performed in collaboration with Dr Neil Berry and Dr Prabha Jayapal at the Dept of Chemistry. A brief description of their methods is given below.

The structural coordinates of NQO1 were retrieved from PDB (1D4A), and modeler 9v was used to generate the model to be used in docking simulations. The quality of the model was assessed using Procheck [187] and Verify3D [188]. Minimization of the model and generation of force fields was performed using GROMACS [189]. Models for the structures of the thiazolides were generated and optimized using Spartan (Wavefunction Inc) (HF/6-31G*). The resulting models were used for docking simulations using Autodock 4.0 [190], a program (14) which is based on a molecular mechanics force field, to model the substrate-enzyme interactions using a genetic algorithm for searching. The simulations were carried out in two stages - "blind docking" and "refined docking". The grid maps representing the protein were calculated with the aid of autogrid, a utility of the autodock software. Two different grid maps with different dimensions ("blind docking" covering the enzyme surface and "refined docking" for the active site) were calculated for the protein. Both in the blind and refined docking, the simulations were carried out using Lamarckian Genetic Algorithm with the initial population of 300 individuals and with the crossover rate, mutation rate and elitism set at the default values of 0.8, 0.2 and 1 respectively. Resulting orientations lying within 2.0 Å in the RMSD were clustered together and represented by the orientation with the most favourable free energy of binding.

Results

To establish and validate the assay, the $K_{m'menadione}$ was sought, for comparison to existing values. In order to do this, a working assay was required

Initial assay conditions

The initial assay conditions, which were based on those suggested by Ernster [185], resulted in an upwards sloping curve (Figure 6.2) that is not characteristic of the expected Michaelis Menten kinetics [191]. Furthermore the duration of the assay was unacceptably long since both NADH and menadione sodium bisulphite are relatively unstable; breakdown of the substrates is likely to be a concern.

Troubleshooting of Assay Conditions

The paragraphs below describe some of the steps that were taken to troubleshoot the unusually slow but accelerating reaction curves.

Initial experiments exhibited low but accelerating rates of reaction (Figure 6.2). Therefore the assay was carried out using various concentrations of enzyme (Figure 6.3) and substrate (Figure 6.4). Alterations in enzyme concentration did not correct the accelerating reaction rate. Furthermore, reactions proceeded in an unexpected order suggesting sources of uncontrolled variation, although initial velocities rose with increasing enzyme concentration (Figure 6.3). Increasing the concentration of MBS resulted in reaction curves in their expected order, with higher substrate concentrations leading to faster

overall rates of reaction although the curves still did not reflect typical enzyme kinetics and saturation of the active site was not achieved. This suggested that MBS concentration might be limiting the reaction rate and it was decided to use both very high concentrations of menadione, but to also to vary the concentration of enzyme over several orders of magnitude. (Figure 6.5). At 1 mM the curve appeared to follow relatively normal enzyme kinetics albeit with a slow overall rate and an initial lag phase. At 50 mM the rate of reaction was unexpectedly slow. This may be due to substrate inhibition which can occur at very high concentrations of menadione. Changing the concentration of enzyme appeared to have no discernable effects on the reaction, which further suggests menadione was the rate limiting factor and that concentrations of enzyme were in excess. To confirm this, the concentration of menadione was added at starting concentrations of 0.1, 1 and 10 mM. 1 U of NQO1 was used. Reactions were performed in duplicate. Figure 6.6 reveals that whilst at 100 μM menadione the unusual upward sloping curve was still present, at 1 mM the reaction appeared more as expected. At 10 mM, unlike in the previous experiment where 50 mM appeared to suppress the reaction, 10 mM menadione resulted in a much quicker reaction rate, consuming over 170 nMol of NADH within the first 5 minutes. Notably, saturation was still not achieved.

In a final attempt to resolve the issues with the assay, the concentration of NADH was increased to 300 μM but NQO1 was reduced to 0.005 U since several previous experiments had indicated that the enzyme may be in excess, and concentrations of menadione were 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, and 60 μM (Figure 6.7). Reactions were performed in duplicates. At 35 μM

menadione, the reaction plateaus after >280 nmoles NADH have been depleted, corresponding to the total amount added to the starting concentration, however curves possess the accelerating form as seen previously were again observed and were also in an unusual order with 25, 35 and 60 μM menadione resulting in the highest reaction rates. Saturation of the enzyme active site was still not achieved.

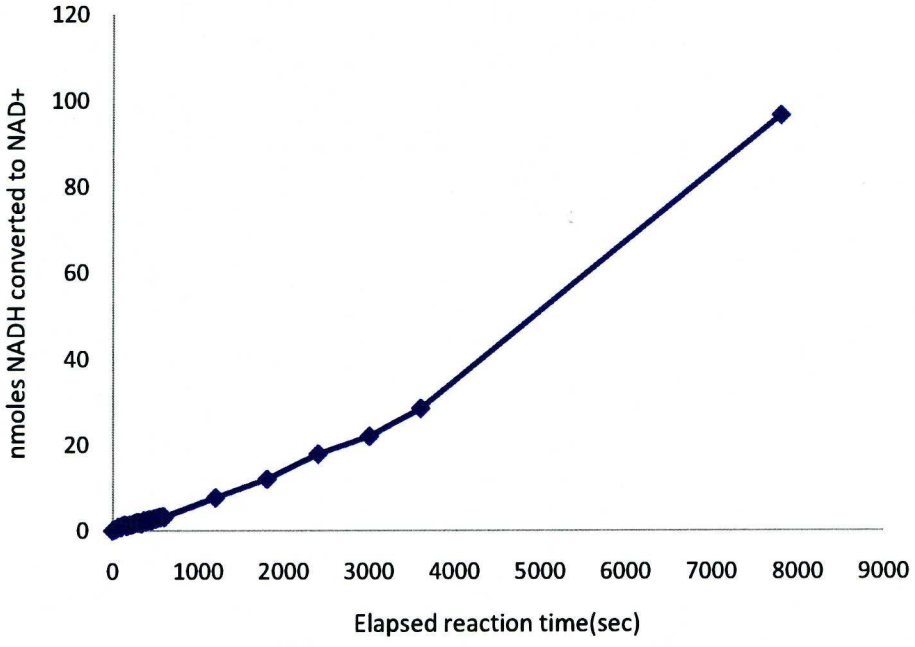


Figure 6.2. Conversion of NADH to NAD+ using 1 unit of recombinant NQO1 in 1mL 50 mM Tris/HCl pH 7.4, 5% DMSO, 200 μM NADH, 65 μM MSB. Reaction was monitored by observing the MSB-mediated conversion of NADH to NAD+ at 340 nm

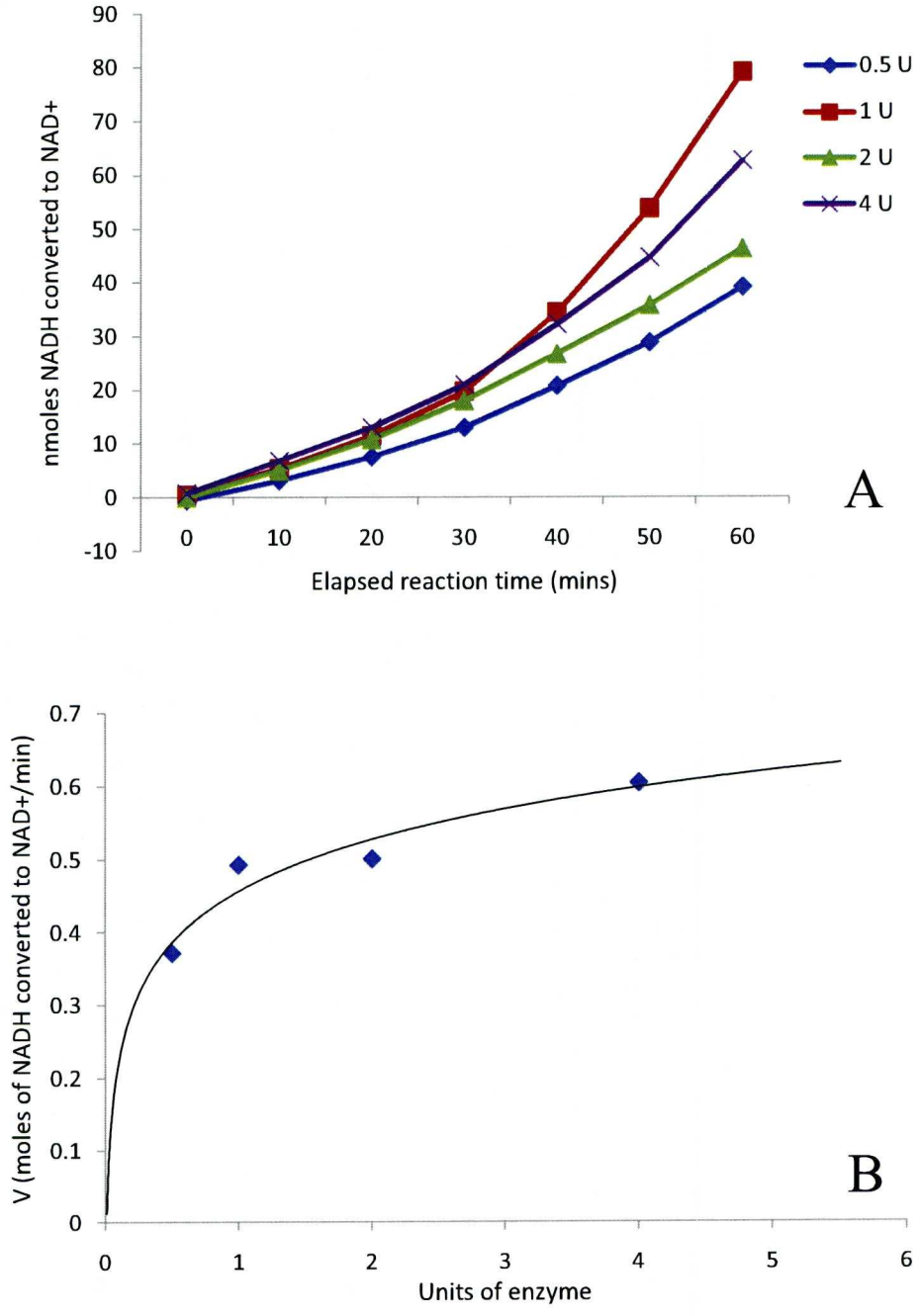


Figure 6.3. Effects of NQO1 concentration on the conversion of NADH to NAD+ by NQO1. Reaction buffer contained 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 200 μM NADH, 65 μM MSB and 0.5, 1, 2 or 4 units of NQO1 per mL. A: Conversion over time of NADH to NAD+; B: Initial reaction velocities.

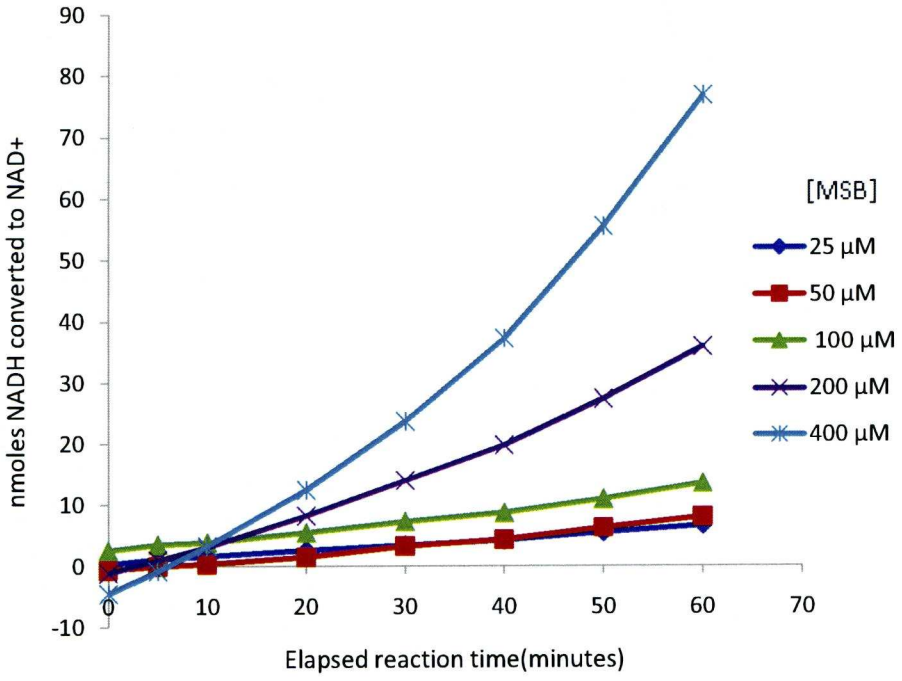


Figure 6.4 Effects of MSB concentration on conversion of NADH to NAD+ by NQO1. Reaction mixture contained 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 200 µM NADH, 1 unit of NQO1/ mL and MSB (25, 50 100, 200, 400 µM). Conversion of NADH to NAD+ was measured at 340 nm. Reactions were performed in singlicate.

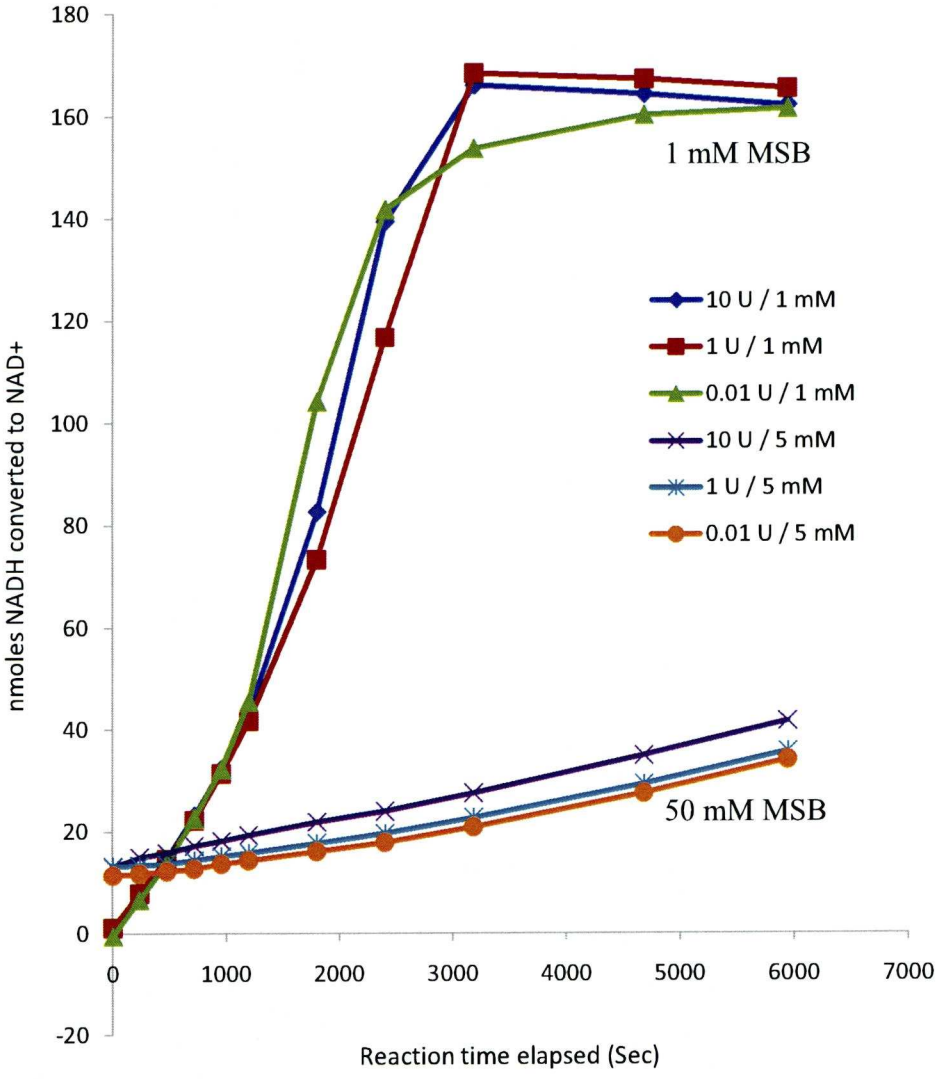


Figure 6.5. Effects of various enzyme concentrations on conversion of NADH to NAD+ by NQO1 at 2 concentrations of MSB. Reaction mixture contained 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 200 μM NADH, MSB (1 and 50 mM), and NQO1 (10, 1 and 0.01 units per ml). Reactions were performed in singlicate.

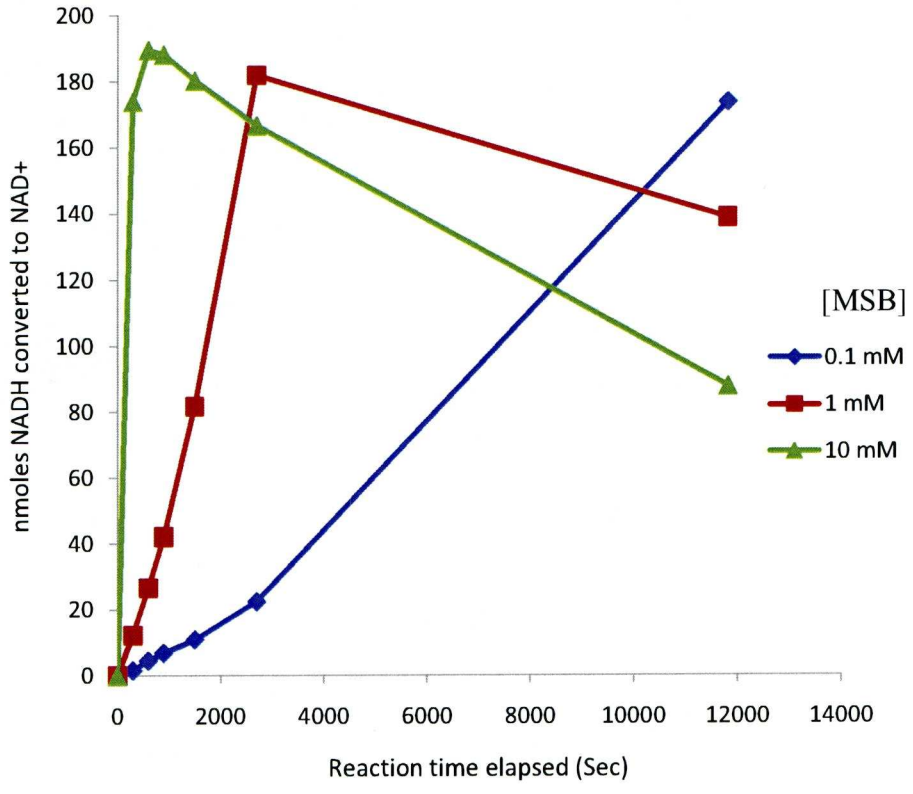


Figure 6.6 Effects of various concentrations of substrate MSB on the conversion of NADH to NAD+ by NQO1. Reaction mixture contained 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 200 μM NADH, MSB (0.1, 1 and 10 mM) and 1 U/mL of NQO1. Despite low concentrations of enzyme, saturation of the active site could not be achieved. Reactions were performed in duplicate.

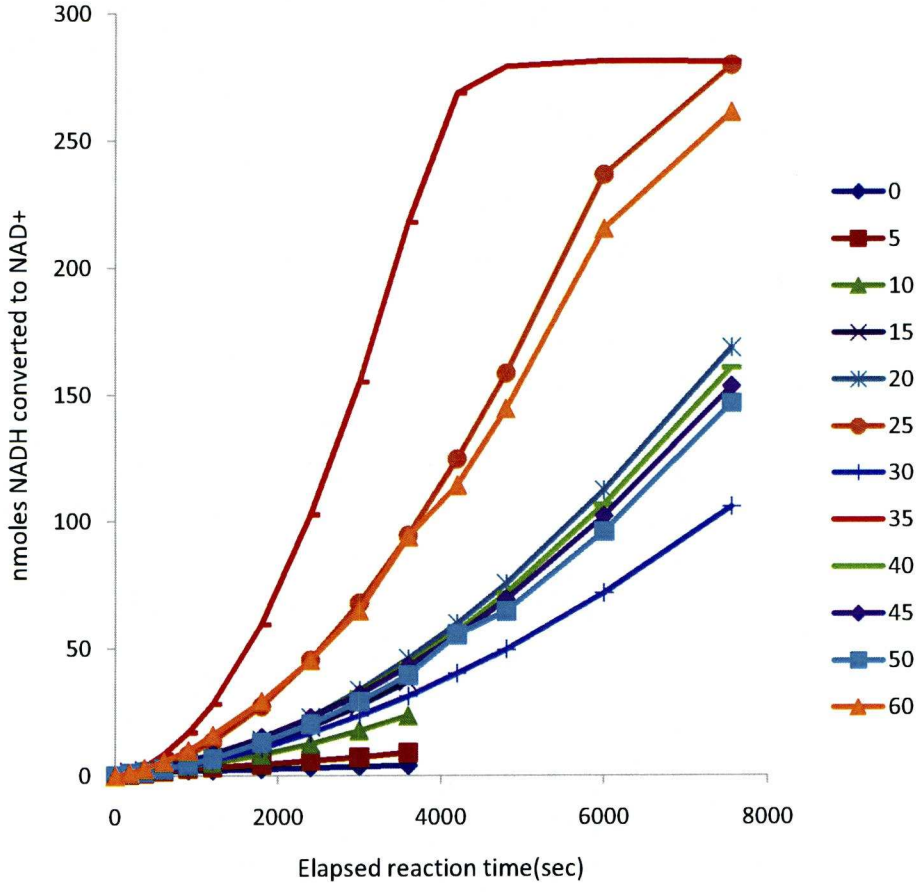


Figure 6.7 Effects of concentration of MSB (0 – 60 μM) on conversion of NADH to NAD+ by NQO1. Reaction mixtures contained 50 mM Tris/HCl (pH 7.4), 5 % (v/v) DMSO, 300 μM NADH, NQO1 (0.005 U/mL) and MSB (0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, and 60 μM).

Monitoring of menadione reduction via reduction of cytochrome C

As earlier attempts to monitor the reduction of menadione using the reduction of NADH at 340 nm had been unsuccessful, a new assay method for measuring the rate of reduction was adapted from Lind *et al.* [186]. After lengthy investigations into what might be causing the problems with the assay, several factors were identified as being potentially problematic and were addressed with a different assay design. The new assay used the menadione mediated reduction of cytochrome C to determine the reaction rate to avoid the problem of overlap of the absorbances of NADH and MBS. Menadione (MEN) was used in place of MBS which should be more heat, light and humidity stable; and several technical improvements were also made to the method to minimise variation between and within experiments. These included starting the reaction by the addition of enzyme and mixing carefully with a clean plastic sparging stick; BSA and Triton-X-100 were also added as activators of NQO1 but DMSO was omitted.

Using 100 ng (0.05 U) and 10 ng (0.005 U) of recombinant human NQO1, reaction kinetics appeared normal (Figure 6.8), with 100 ng proceeding more quickly than with 10 ng and reaching a plateau as would be expected

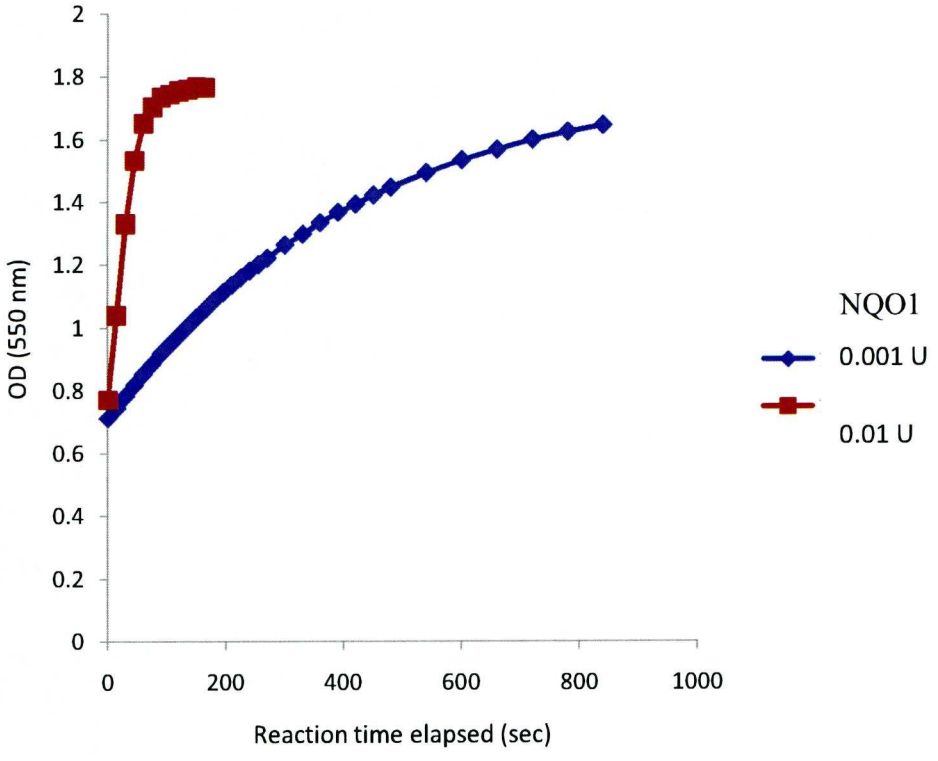


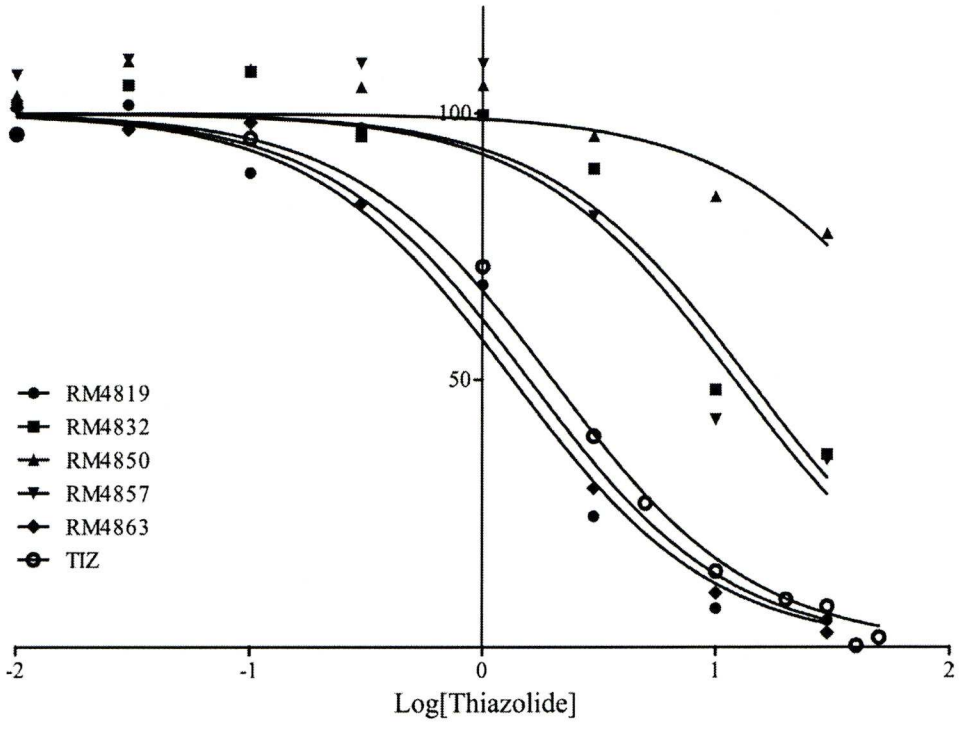
Figure 6.8 Reaction kinetics of MEN-mediated reduction of cytochrome C at 550 nm at 2 concentrations of NQO1 (0.001 and 0.01 units/mL). Reaction mixture contained 10 μM MEN, 50 mM Tris/HCl (pH 7.4), 500 μM NADH, 0.08 % (v/v) Triton-X-, 0.07 % (w/v) bovine serum albumin and 77 μM cytochrome C in a final reaction volume of 1 mL.

$K_{m,menadione}$ Determination

With the parameters of the assay apparently within normal limits, the suitability of the assay system for probing the interaction with the thiazolides was determined by measuring the $K_{m'menadione}$. The value was calculated in Prism (Graphpad) using non-linear regression, fitting a classical Michaelis Menten kinetic model, yielding a value of 2.45 μM (SE ± 0.34). Later whilst inhibition type was being determined, the non-linear regression was used again to fit new data to an uncompetitive model of inhibition, yielding a new value of 2.15 μM (SE ± 0.32)

TIZ and five other thiazolides inhibit NQO1.

The interaction between NQO1 and thiazolides was investigated further by assaying the reduction of menadione to menadiol, in the presence of thiazolides. All of the compounds tested exhibited an inhibitory effect on NQO1 under the conditions used (Figure 6.9). TIZ, RM4819 and RM4863 showed IC50 values in the low micromolar range (~3 μM) whilst RM4832 and RM4857 had higher IC50 values of 13.92 μM and 12.04 μM, respectively. However, some low-level inhibition of NQO1 was observed with RM4850, the relative insolubility of the thiazolides meant that an IC50 value could not be accurately reported in this case.



Compound	IC50 (μM)	95 % CI	Goodness of fit - R^2
TIZ	2.00	1.73 - 2.31	1.00
RM4819	1.59	0.99 - 2.55	0.97
RM4832	13.92	10.18 - 19.03	0.94
RM4850	>30	-	-
RM4857	12.04	6.85 - 21.17	0.89
RM4863	1.35	1.09 - 1.68	0.96

Figure 6.9. Concentrations of 7 thiazolides causing 50 % inhibition (IC50) of NQO1 activity. Using a range of concentrations of the thiazolides, IC50 values for the inhibition of NQO1 were determined. Curves were fitted to the data points using non-linear regression, with PRISM (GraphPad).

Determination of inhibition type

To further investigate the interaction of TIZ with NQO1 and to elucidate further information regarding the type of inhibition, different concentrations of MEN (1, 2, 4, 6, 10 μM) were varied over different [TIZ] (0, 0.5, 1, 2, 10 μM). Initial reaction velocities were recorded and plotted on a double reciprocal (lineweaver-burke) plot of 1/[menadione] vs. 1/initial velocity (Figure 6.10). With the exception of measurements taken at 0 μM concentration of TIZ, which deviated slightly, lines of best fit to the data lie parallel to each other, a characteristic seen only in uncompetitive inhibition[191]. This was confirmed using GraphPad Prism to fit the data to uncompetitive (inhibitor binds substrate-enzyme complex), non-competitive (inhibitor acts without affecting substrate binding) and competitive (substrate and inhibitor compete for binding) inhibition models. Examination of the R^2 values and residual plots for the fitted models confirmed that uncompetitive inhibition best described the data (not shown).

In silico modelling of the interaction of the thiazolides and NQO1.

In parallel to the *in vitro* data, to investigate how and where the thiazolides under consideration may bind to NQO1, a molecular modelling docking experiment was performed in collaboration with Prabha Jayapal of Dept. of Chemistry. An *in silico* model of NQO1 was built based on the previously derived X-ray crystal structure (PDB: 1D4A) [178]. Docking simulations were performed using Autodock 4.0 to identify a likely binding site, and binding modes of the thiazolides. "Blind docking" experiments revealed that the most

favourable binding energy was observed for nitazoxanide in the pocket close to the FAD co-factor. Refined docking experiments were performed in this region using all the molecules. The best calculated free energy of binding (ΔG_b) for each thiazolide is shown in Table 6.1. An example of a conformer from the lowest energy cluster of each docking simulation is shown in Figure 6.11. Comparison with the crystal structure of NQO1in complex with the uncompetitive inhibitor dicoumarol (Figure 6.12) indicated that the location for TIZ binding is the same. The accuracy of the model is supported by the experimental results suggesting that TIZ is also an uncompetitive inhibitor of NQO1.

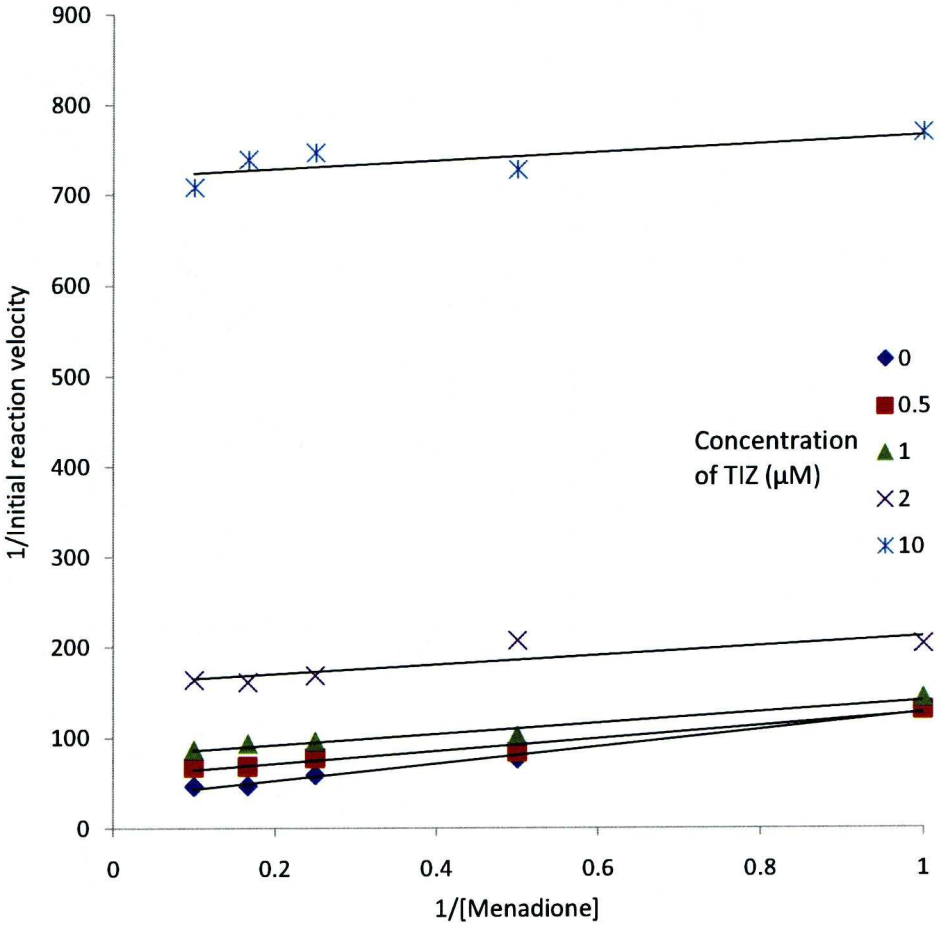


Figure 6.10. Effects of changing the concentration of MEN on the reduction of cytochrome C at 5 fixed concentrations of TIZ. The double reciprocal plot shows parallel lines characteristic of uncompetitive inhibition.

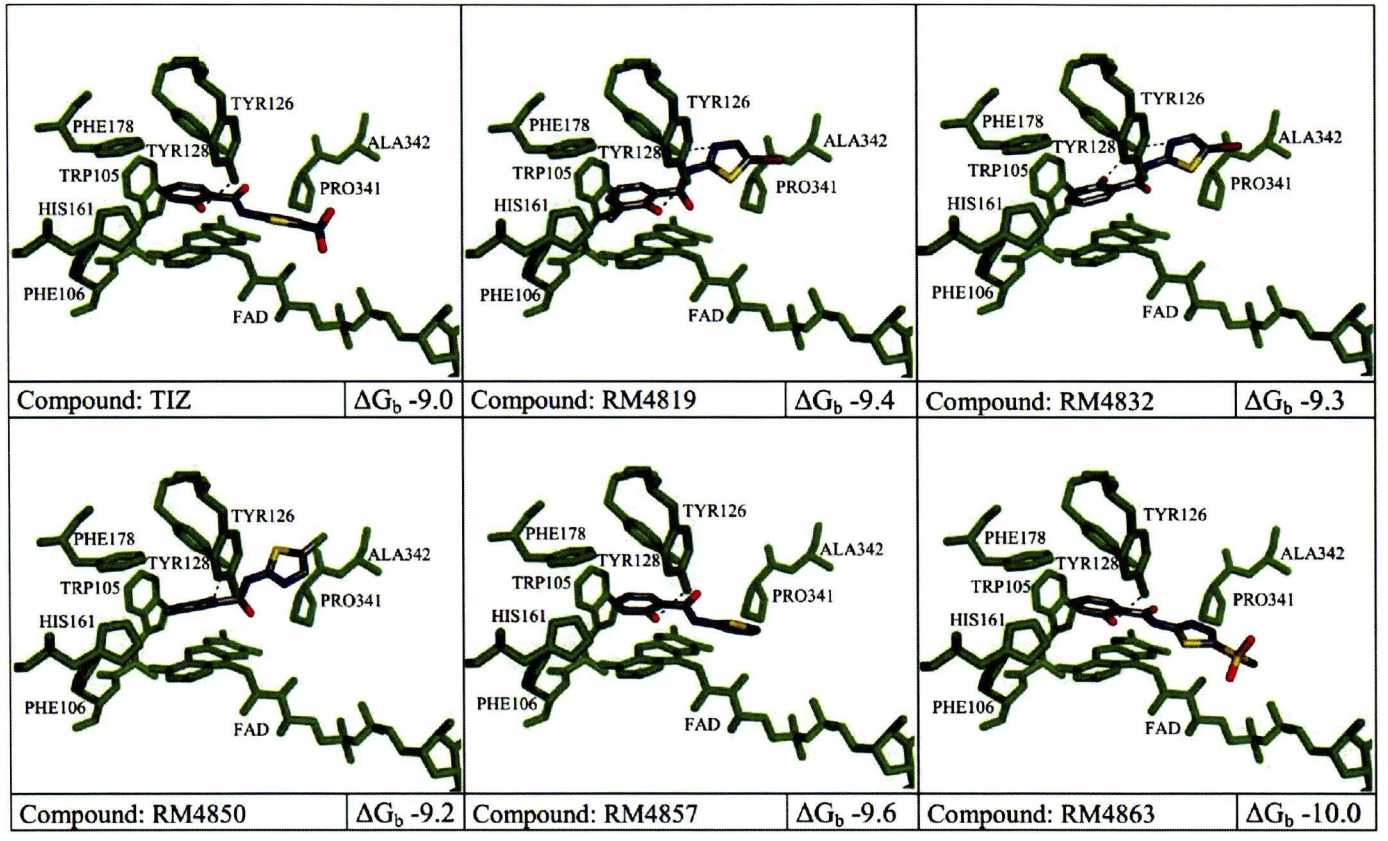


Figure 6.11 Results of docking simulations for the thiazolides. An example of a conformer from the lowest enery cluster is shown for each thiazolide

Compound	ΔG_b
TIZ	-9.0
RM4819	-9.4
RM4832	-9.3
RM4850	-9.2
RM4857	-9.6
RM4863	-10.0

Table 6.1. *In silico* determinations of free energy binding (ΔG_b) of TIZ and its analogues.

Discussion

Using affinity chromatography, a binding interaction between human NQO1 and an immobilized conjugate of TIZ was identified. To determine the physiological significance of this interaction, enzyme kinetic studies were performed using TIZ and related thiazolides.

Initial kinetic experiments using NADH as an indicator of quinone reduction failed to yield useful results, regularly producing reaction rate curves that were problematically slow, which suggested that there were also issues affecting reproducibility. Whilst the latter was easily addressed with methodological changes, any single cause of the slow yet accelerating reaction rates remains somewhat elusive, but several possible explanations for the accelerating rates are proposed. With respect to MBS, three areas for concern were identified. Firstly, at high concentrations, MBS is absorbed at 340 nm. Secondly, high menadione concentrations can also inhibit the enzymatic activity of NQO1. Finally, menadione sodium bisulphite, chosen for its water solubility, has limited stability to heat, light and humidity. Therefore, over time, the breakdown of MBS, the relief of substrate inhibition and concomitant decrease in OD as MBS was reduced by NQO1, could have all contributed to the apparent accelerating rate of the reaction curves. Breakdown of NADH may have also exacerbated this. Breakdown of MBS and NADH would not be a problem over a short assay time; however the very slow rates of reaction meant that the assay times were extended to obtain a meaningful change in OD. At this point it is worth noting that the concentrations of MBS that were used in several of the troubleshooting experiments were far in excess of the k_m

and should have been at saturating concentrations, this suggests that the enzyme concentration was in great excess.

Looking retrospectively at the troubleshooting experiments where the enzyme concentration was varied, a plot of enzyme concentration against initial velocity (Figure 6.3, lower panel) reveals downward curving dependency of initial velocity on enzyme concentration. Tipton [191] describes one of the common causes for this can be contamination of the enzyme solution with dissociable inhibitor. Without rigorous testing we cannot be certain if the enzyme solution was contaminated with inhibitor. However, according to the manufacturer's specifications, the specific activity of the first vial (263 U/mg) was nearly half of that of the new batch (492 U/mg). Considering everything that is now known about the initial attempts to establish a working assay, it is still difficult to determine the precise cause of the problems although several factors indicate that the problem was probably associated with the enzyme. The initial reaction conditions were based upon a method widely used in studies of NQO1 [185] and several publications have used the method successfully [192].

Using the new assay, the $K_{m,menadione}$, was measured to be 2.5 (±0.4) μM . This is in good agreement with previously published values of between ≈ 2 and 4 μM [193, 194]. Initial single point experiments (not shown) indicated TIZ may be an inhibitor of NQO1 so IC50 values for the various thiazolide compounds were obtained. All of the compounds tested exhibited some inhibitory activity against NQO1, although for some of the compounds an IC50 value could not be accurately determined due to the limits of solubility in this particular assay

system. The experimental data yielded low μM (<3 μM) IC50 values for TIZ, RM4819 and RM4863 that would be easily physiologically achievable (21), whilst slightly higher IC50 values were obtained for RM4832, and RM4857 (14 and 12 μM respectively). RM4850 had a much lower, but still significant inhibitory effect on NQO1 (>30 μM). The results suggest that TIZ is a uncompetitive inhibitor relative to menadione. An uncompetitive inhibitor binds the enzyme-substrate complex, however in ping pong kinetic enzymes this result can arise if the inhibitor is competitive with respect to the first substrate, in this case NADH. The docking simulations indicate that TIZ and the other thiazolides bind in the same way as dicoumarol, and inhibitor of NQO1. Dicoumarol competes with NADH for access to the active site of NQO1, but is uncompetitive with respect to its quinone substrates, supporting the accuracy of the docking result.

The calculated ΔG_b values from the docking simulations all show similar, relatively strong binding of all thiazolides studied but no significant differences in the predicted binding energies between the molecules studied were found. The *in silico* calculations indicate that the thiazolides bind to the active site and may share many of the same interactions, but there were discrepancies between the ΔG_b *in silico* data and the IC50 *in vitro* data. Certain assumptions about the state and flexibility of the active site must be made before docking calculations can proceed, which is an inherent limitation of computational modelling. However, the modelling has relevance to affinity chromatography, and an examination of the docking simulations reveals several common features that could account for the observed binding affinities.

The importance of the phenyl ring in affecting thiazolides binding to NQO1, as implicated by the affinity chromatography, is further emphasized by the docking simulations. For all molecules there was predicted to be a co-planar π-stacking interaction (~3 Å separation) between the phenolic portion of the compounds with the ring structure of FAD. All molecules also display an edge-to-face π-stack with TRP105. The binding pocket around the head group of FAD is aromatic rich, thus the possibility for π-stacking and hydrophobic contacts seems a likely driving force for binding. The additional methyl group on the phenyl ring of RM4819 may also contribute an additional favourable hydrophobic interaction compared with the other molecules under investigation. Many of the docking simulations revealed the potential for the formation of hydrogen bonds with either TYR126 or TYR128 (as indicated by dashed black lines in Figure 6.12) although the exact position of these residues *in vivo* is uncertain because the residues are mobile to some extent, forming a 'lid' to the active site. The high degree of similarity between these interactions and those between NQO1 and its natural substrates indicates further that this is a highly viable possibility for the location of binding between TIZ and the active site of NQO1

TIZ, RM4832, RM4850 and RM4863 have been tested for their potency against HCV and HBV replication [13] and the IC50 results for these compounds correlate very well with the data acquired for the potency of the thiazolides against HCV replication, and share the same order of potency. It is unclear at this point if this correlation is merely coincidental or if there is some

other connection between NQO1 and HCV infection, although one candidate for the link is the tumor suppressor protein p53 which is stabilized by NQO1 [180], and inactivated by the non-structural protein 5 A (NS5A) of HCV [195]. It is possible that the interaction of thiazolide and NQO1 causes a favourable change in conformation that promotes stabilization and accumulation of p53, or that a transient increase in oxidative stress due to NQO1 inhibition, stimulates a new round of anti-oxidant response element controlled genes such as nqo1, which could also lead to p53 stabilisation. Stabilisation of p53 could also be a candidate for the cell cycle-dependent, anti-proliferative effects of the thiazolides observed by Muller *et al.* [196]. Furthermore, p53 has also been shown to mediate apoptosis upon chronic activation of the unfolded protein response (UPR) [197] an integral process to the pathogenesis of HCV [198]. The UPR is triggered by an accumulation of unfolded protein (i.e. during viral infection) in the lumen of the endoplasmic reticulum causing BiP, which acts as a master control switch of the UPR, to dissociate from proteins ATF6, IRE1 and PERK which subsequently become active and lead to protein folding, protein degradation and translational attenuation. In particular, translational attenuation results from the phosphorylation of eIF2α. Some proteins of HCV have been shown to inhibit PERK, thus increasing the overall levels of protein synthesis.

Based soley on the observation that thiazolides inhibit NQO1, it is difficult, if not impossible, to conclude that NQO1 represents a host-based mode-of-action that would explain NTZs wide-spectrum of effects. However, as the latest in an increasingly long list of putative targets, the possibility of multiple modes

of action becomes a probability, and the likelihood that, *in vivo*, the possibility that NTZ targets or interacts with only a single enzyme seems less likely. In a study to determine possible HCV resistance to thiazolides, Korba *et al.* [12] hypothesize that these compounds can mediate their anti-viral effects through the host cell. The results of the present study appear to support this hypothesis and demonstrated direct interactions between the thiazolides and host cells.

Having identified another putative target of the thiazolides and also investigated proteomic and transcriptomic changes modulated by thiazolides in two differing cellular systems, the question of whether the results of these experiments relate to one another remains. The final chapter of this thesis will discuss the findings of the present study and the research of others and see if any further light can be shed on the mode of action of the thiazolides.

Chapter 7

General Discussion

Nitazoxanide seemingly provides potential therapeutic solutions to several problems of global clinical significance to humans. The range of maladies against which NTZ has a demonstrated potential is remarkable and includes, but is not limited to, diseases caused by anaerobic bacteria, parasites such as *Giardia lamblia* [199] and *Cryptosporidium parvum* [200], several viruses including HCV, HBV [13], Rotavirus [20, 21] and Influenza A virus [201] and even the proliferation Caco2 colon carcinoma cells [166]. This wide range of activity suggests that NTZ may have multiple modes of action; the discovery of multiple putative host and pathogen targets supports this, as do the accounts of both its activity against axenic *G. lamblia* cultures and host mediated resistance mechanisms in viral replicon cells.

The primary aim of the present study was to investigate the mechanisms of action of NTZ against a parasite and a virus. In Chapter 3, the differences in protein expression between a NTZ resistant strain of the intestinal parasite *Giardia lamblia* and a wild-type strain were investigated using DIGE. The protein changes observed in this experiment were also quantified on a transcriptional level using RT-qPCR and in addition to this, a microarray comparison of a resistant strain was reanalysed to gain new perspectives on global transcriptomic changes. A DIGE analysis of a NTZs/METr strain of *G. lamblia* was also analysed to address the issue of cross resistance of the NTZr strain to MET. The implications of these results were considered on the basis of

both individual proteins, by grouping the types of proteins involved and also, where possible, using bioinformatic techniques to describe the changes in functions and processes that were unique to the NTZr strain. The data indicated that a group of annexin repeat proteins named 'protein 21.1' with various predicted functions were associated with NTZ resistance in *G. lamblia.* Unfortunately virtually nothing is known about these proteins and therefore without further study, little can be concluded regarding their functions and role in resistance. The large number of differentially expressed VSP genes and modulations in proteins related to cell mitosis suggested that cell cycle events may be affected by NTZ. Comparison of molecular functions and biological processes using GO terms indicated that amino acid synthesis, translational elongation, protein folding and unfolded protein binding may also be important processes modulated by NTZ. The bioinformatic interpretation of the results in Chapter 3 is likely to improve with further refinements to the gene models and more comprehensive annotation of the genome.

In Chapter 4, the DIGE technique was used to study the effects of NTZ on HCV, by using an Huh7 cell line expressing a sub-genomic replicon of HCV treated with thiazolide. The replicon lacks the envelope and core proteins of the full virus and thus cannot form infectious viral particles, but is still able to replicate the other viral proteins and is a good model for HCV replication overall. Large changes in heat shock proteins, including BiP, a protein that is critical in regulating the UPR [136], and host elements with importance to HCV replication complexes [159] were evident and were confirmed by bioinformatic

analysis. Several nucleotide binding proteins and proteins with glycolytic functions were also differentially expressed. The high quality functional annotation data for human proteins (in contrast to that for *G. lamblia*) was used with three bioinformatic techniques to investigate the data. The greatest amount of information was gained from PantherDB [151] and DAVID [152, 153, 202] which use various gene annotations to describe the data and quickly confirmed similar areas of significance as described above. Interactions between the thiazolide modulated proteins were also investigated using STRING [154]. The potential of this tool for discovering groups of proteins within datasets that might be functionally related was illustrated by its being the only technique to isolate the group of proteins with glycolytic associations, but the reliability of the interactions was somewhat undermined by the possible presence of errors in the network due to automatically text-mined interactions.

After looking at changes in the global expression in systems of importance to host and pathogen interactions, affinity chromatography was used to establish interactions between host proteins and thiazolides. Non-specific binding of proteins to the resin and spacer arm initially proved problematic. However by using a TIZ-agarose conjugate, a protein of approximately 31 kDa was repeatedly eluted with 1 mM TIZ buffer. This protein was not eluted when thiazolide derivatives were conjugated to the agarose resin in an alternative orientation. The protein eluted with TIZ was identified as NAD(P)H Quinone Oxidoreductase 1, a FAD binding protein that catalyses the 2 electron reduction of quinones to their less toxic hydroquinone form via mechanisms that do not generate toxic radical species [182]. In Chapter 5, a detailed study on the

interaction between NQO1 and thiazolides showed that NQO1 could be inhibited by TIZ and several other thiazolides. The inhibition type was also determined to be uncompetitive with respect to menadione, binding the enzyme substrate complex. The results of the docking studies indicate that TIZ binds the same location and interacts with the active site in the same manner as other inhibitors, such as Dicoumarol [203] which is also uncompetitive with respect to menadione, supporting the accuracy of the model; however the ΔG_b values did not correspond to the measured IC50 values.

The most definitive finding of this study was the inhibitory activity of the thiazolides against NQO1. It is unclear how this fits into the wider picture, and the question of whether the inhibition of NQO1 could cause some or all of the proteomic changes that have been observed in RP7 cell lines requires further study before this question can be conclusively answered. The effects of dicoumarol, which is an inhibitor of NQO1, on the proteome of Huh7 cells and on HCV's ability to replicate, should provide answers to this question. Both the experimental and modelling data support the notion that thiazolide binding to NQO1 occurs in a manner analogous to other inhibitors such as dicoumarol. NQO1 has been shown to be up-regulated in certain types of tumour cell lines and has been targeted using 'pro-drugs' which become toxic to the cell upon activation by NQO1 [204]. TIZ has also been shown to inhibit proliferation of rapidly proliferating Caco2 cells, a colorectal adenocarcenoma cell line, although in the same study it was hypothesised that this activity was mediated through GSTP1 [166]. Nonetheless, the inhibition of NQO1 could be an important mechanism for the survival of cells in which NQO1 is over-expressed.

NQO1 also has a role in stabilising the tumour suppressor protein p53, which controls cell cycle progression and DNA repair [180, 192]. Simultaneous interference with this interaction, as occurs with dicoumarol, in combination with the inhibition of NQO1, which is an important detoxifying enzyme, could leave cells vulnerable to the deleterious effects of oxidative stress arising from defects in protein folding.

Based on the results of the DIGE experiments, it has not been possible to draw a single definitive conclusion about the mechanism of action of thiazolides. Most proteomic techniques are probably best considered to be hypothesis generating however, since their results alone cannot often provide direct evidence for a particular interaction, The DIGE experiments have, however, suggested that there may be several biological areas which are of particular importance to the mechanism of action of thiazolides. Despite fundamental differences between the proposed mechanisms of action of thiazolides against *G. lamblia* and HCV, the results of the two DIGE experiments share some functional themes. These functional overlaps are primarily protein folding and unfolded protein binding and the data suggests that HSPs may have significant roles in both systems. Whether this is the result of common systemic effects of the drug that do not represent the primary mode of action i.e. drug stress, or represents a previously undiscovered common contribution towards the mechanism of action in both systems is unclear. What *is* clear is that this result warrants further study in both systems and may be of particular relevance to the HCV replicon system where

ER stress response mechanisms are integral to viral infections [142, 144, 146, 155].

There are still some fundamental issues that need to be addressed concerning potential resistance mechanisms of *G. lamblia* that might not be detected by either DIGE or by the microarray experiments used in this study. A closer look at the underlying genetic sequences of resistant pathogen strains can in many cases reveal mutations that are responsible for resistance. A recent study using high resolution tiling arrays to discover resistance conferring mutations to MET in *H. pylori* [100] is a good example of this. Whole transcriptome shotgun sequencing (WTSS) techniques such as RNA-seq which provide both transcript sequence data and transcript abundance could now also be used to improve on the tiling array method and identify genes that might carry mutations. However since resistance to NTZ can be lost after a round of en- and excystation, the mechanism of resistance may be mediated by an epigenetic mechanism rather than permanent genetic change such as a mutation. Techniques such as ChIPseq, which is a combination of chromatin immunoprecipitation of proteins suspected of interactions with DNA combined with sequencing of the DNA fragments to which they bind, could also have a place in determining if epigenetic mechanisms are involved in resistance. Other epigenetic mechanisms such as PTMs (histone acetylation for example) could also be investigated using 2DE, and MS. With this data in hand, a more complete picture of how the thiazolides affect cells on a global scale could be built. Sequence analysis of a recently created HCV replicon cell line, resistant to NTZ through what are believed to be host cell mediated means [12], could also be beneficial to the discovery of NTZs

targets. Future studies of the mechanism of action of the thiazolides against HCV should also consider using genomic length replicon expression systems that include the structural proteins of HCV [205]. These newer replicon systems that include HCV glycoproteins E1 and E2 might be especially important in light of evidence that NTZ affects the maturation of Influenza A, viral haemmaglutinin [201] and the effects of glycoprotein maturation in Rotavirus is also being investigated. The process of glycoprotein maturation is an ER initiated process that continues into the golgi, again highlighting the importance of protein metabolism as a potential target of action of the thiazolides

There were some areas of the present study, which were ultimately hindered by limitations currently imposed on them. The bioinformatic aspects of studying *G. lamblia* presented considerable challenges. After identifying proteins that were modulated in NTZ resistant *G. lamblia*, functional annotations for these proteins were sought. The only functional annotation currently available for *G. lamblia* are GO terms, and only 25 % of proteins in the 1.2 release of GiardiaDB have GO annotations assigned to them. Even for relatively large datasets such as microarrays, reducing the number of significant results by this proportion leads to rather small sample sizes for hypergeometric overrepresentation analysis. Furthermore, obtaining the mappings of GO terms to the GiardiaDB identifiers and transcribing them into a form useable by other tools was a difficult task, requiring at least some basic knowledge of regular expressions (a method commonly found in text editors or programming languages for of identifying characters, words, or patterns of characters etc). This is in contrast to other organisms such as *Saccharomyces cerevisiae* for which the information is

readily available. Although there are now many tools available for GO term enrichment analysis, these tools often exclude all but the most common species. *G. lamblia* was even excluded from DAVID, despite obtaining its information from several databases that do contain gene information for this species, such as NCBI (although the current private beta release addresses this [202]).

Whilst microarrays have a clearly defined background population set and thus are highly amenable to statistical analysis, most proteomic techniques suffer from an unknown background population. The main reason for this, particularly in parasitic organisms but that is likely true of any organism that can survive in more than one environment, is the potential for unknown differences in the expressed proteome that is subject to fluctuations from cellular differentiation and life cycle stages [206, 207]. A method to account for post-translational modification (PTMs) and alternative splicing also needs to be factored into statistical models before they can be considered accurate. One way to solve this problem whilst taking into account the technical biases of DIGE (i.e. protein solubility and dynamic range in a gel based system) might be to pick and identify every detectable protein on the analytical DIGE gel, thus quantifying the background against which, statistical analysis of the proteins can be performed. The time required to identify the several thousand resolvable protein spots that are detectable on the gel however means this is somewhat impractical. Gel-free techniques such as multidimensional protein identification technology (MudPIT) that can identify hundreds of proteins in a sample [208, 209] could be used to achieve similar statistical gains, however this would not incorporate

some of the advantages of gel-based methods such as the ability to recognise phosphorylation simply [210].

Perhaps the best approach to the statistical challenges that are faced by most current proteomic techniques is not to develop new statistical analytical methods, but to create or adapt methods that can be analysed using the existing robust methods. There are some emerging proteomic techniques that do not suffer from the statistical challenges that techniques such as DIGE do; protein arrays use a defined set of proteins as the background population, and thus eliminate the uncertainty in the statistical analysis. Two methods in particular stand out as promising techniques for the discovery of drug mechanism of action. Global effects of drugs on the protein expression could be quantified by using antibody arrays (or capture arrays) that possess specificity for individual proteins. By detecting the concentration of protein bound by each spot, the concentration can easily be determined. The bottle-neck to this is clearly the availability of specific antibodies for every protein of interest. However by using the results of global proteomic techniques, such as DIGE, iTRAQ, SILAC or ICAT in combination with protein-protein interaction resources such as Reactome, Biogrid or STRING, protein arrays that are targeted to sub-proteomes and would benefit from straight forward statistical analysis could be created. A functional protein array meanwhile, with proteins immobilised to the array, could be used to directly observe which proteins the drugs bind to. Protein synthesis and capture is the main obstacle to this, but improvements in recombinant expression have led to the ability to perform *in situ* synthesis of proteins directly on the array. Nucleic acid programmable protein arrays

(NAPPA) [211] are an example of this, as is synthesis of proteins directly from a DNA microarray, "DNA to protein array" (DAPA) [212] for example. The implications of these techniques to the field of drug target discovery could be immense and in combination with techniques such as ChIP on chip and CHIPseq which can determine the interactions between DNA and proteins, the improvements in dynamic range, resolution, reliability and interoperability of the techniques takes the field a step closer to realising the goal of global systems biology. Increases in the capacity of techniques such as multiple reaction monitoring (MRM) [213], a mass spectrometric technique which use single peptides as representatives of protein concentration, is also helping to establish this technique as a useful method for multiplexing the quantitation of individual proteins and their turnover.

The pull down assay was confirmed as a good way of identifying interactions but establishing the nature of the interaction took several weeks. With higher throughput methods that could result in many candidate proteins, expressing the recombinant protein and testing the interactions with a target would be very lengthy and laborious. There may be ways of prioritising the analyses of these interactions. Protein arrays in combination with biosensing techniques such as dual polarisation interferometry (DPI) [214, 215] could potentially be used to rapidly assess the binding of small molecules to many different proteins without the need for tagging and labelling strategies. DPI uses the principles of surface plasmon resonance to detect distortions in evanescent waves that results from changes in the conformation of molecules at the surface of chip. The technical

limitations of this are unknown, but could be a fast way of finding the targets of small molecules.

Using proteomic, transcriptomic, bioinformatic and classical biochemical and enzyme kinetic techniques, the present study has identified areas for further research in the discovery of the mechanism of action of the thiazolides. Testable hypotheses can now be generated on the basis of the 'omic' results for both *G. lamblia* and HCV and the relative contribution towards the mechanism of action that each of the putative areas of significance makes can be assessed. Most significantly, this study has identified a novel target of the thiazolides, NQO1 has been identified; the importance of this interaction should be assessed by further studies. There are however still several, yet-to-be investigated areas, such as the potential for genetic mutations and how epigenetic mechanisms might contribute to mechanism of action that must be also be addressed if we are to completely understand the complex biochemistry that underpins the activity of the thiazolides.

References

1. Euzeby, J., S. Promtep, and J.F. Rossignol, *Experimentation of Nitazoxanide Anthelmintic Properties on Dog, Cat and Sheep.* Revue De Medecine Veterinaire, 1980. **131**(10): p. 687-&.

2. Rossignol, J.F. and H. Maisonneuve, *Nitazoxanide in the Treatment of Taenia-Saginata and Hymenolepis-Nana Infections.* American Journal of Tropical Medicine and Hygiene, 1984. **33**(3): p. 511-512.

3. Megraud, F., et al., *Nitazoxanide. A possible alternative to metronidazole in the regimens aiming to eradicate Helicobacter pylori.* Gut, 1997. **41**: p. A100-a101.

4. Dubreuil, L., et al., *In vitro evaluation of activities of nitazoxanide and tizoxanide against anaerobes and aerobic organisms.* Antimicrobial Agents and Chemotherapy, 1996. **40**(10): p. 2266-2270.

5. Doumbo, O., et al., *Nitazoxanide in the treatment of cryptosporidial diarrhea and other intestinal parasitic infections associated with acquired immunodeficiency syndrome in tropical Africa.* American Journal of Tropical Medicine and Hygiene, 1997. **56**(6): p. 637-639.

6. Cabello, R.R., et al., *Nitazoxanide for the treatment of intestinal protozoan and helminthic infections in Mexico.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1997. **91**(6): p. 701-703.

7. Adagu, I.S., et al., *In vitro activity of nitazoxanide and related compounds against isolates of Giardia intestinalis, Entamoeba histolytica and Trichomonas vaginalis.* Journal of Antimicrobial Chemotherapy, 2002. **49**(1): p. 103-111.

8. Hemphill, A., J. Mueller, and M. Esposito, *Nitazoxanide, a broad-spectrum thiazolide anti-infective agent for the treatment of gastrointestinal infections.* Expert Opinion on Pharmacotherapy, 2006. 7(7): p. 953-964.

9. Rossignol, J.F., B.E. Korba, and S.M. Kabil, *Nitazoxanide in treating chronic hepatitis C: in vitro activity and a clinical case report.* Gastroenterology, 2006. **130**(4): p. A841-a841.

10. Rossignol, J.F., B.E. Korba, and S.M. Kabil, *Nitazoxanide in treating chronic hepatitis B: in vitro activity and a clinical case report.* Gastroenterology, 2006. **130**(4): p. A848-a848.

11. Kolozsi, W., et al., *Treatment of chronic hepatitis B (CHB) with nitazoxanide (Ntz) alone or ntz plus adefovir (ADV) for two years with loss of hepatitis B e antigen (HBeAg) and hepatitis B surface antigen*

(HbsAg): Report of two cases. American Journal of Gastroenterology, 2008. **103**: p. S150-S151.

12. Korba, B.E., et al., *Potential for Hepatitis C Virus Resistance to Nitazoxanide or Tizoxanide.* Antimicrobial Agents and Chemotherapy, 2008. **52**(11): p. 4069-4071.

13. Korba, B.E., et al., *Nitazoxanide, tizoxanide and other thiazolides are potent inhibitors of hepatitis B virus and hepatitis C virus replication.* Antiviral Research, 2008. **77**(1): p. 56-63.

14. Rossignol, J.F., et al., *Randomized controlled trial of nitazoxanide-peginterferon-ribavirin, nitazoxanide-peginterferon and peginterferon-ribavirin in the treatment of patients with chronic hepatitis C genotype 4.* Journal of Hepatology, 2008. **48**: p. S30-S30.

15. Rossignol, J.F., A. Elfert, and E.B. Keeffe, *Evaluation of a 4 Week Lead-in Phase with Nitazoxanide (Ntz) Prior to Peginterferon (Pegifn) Plus Ntz for Treatment of Chronic Hepatitis C: Final Report.* Hepatology, 2008. **48**(4): p. 344a-345a.

16. Rossignol, J.F., et al., *Clinical trial: randomized, double-blind, placebo-controlled study of nitazoxanide monotherapy for the treatment of patients with chronic hepatitis C genotype 4.* Alimentary Pharmacology & Therapeutics, 2008. **28**(5): p. 574-580.

17. Rossignol, J.F., et al., *Randomized double blind placebo-controlled trial of nitazoxanide in the treatment of patients with chronic hepatitis C genotype 4.* Journal of Hepatology, 2008. **48**: p. S311-S312.

18. Schaninger, T., J. Hong, and G.G. Luo, *Nitazoxanide Inhibits Hepatitis C Virus Replication in Vitro.* Hepatology, 2008. **48**(4): p. 756a-756a.

19. Lanata, C.F. and M. Franco, *Nitazoxanide for rotavirus diarrhoea?* Lancet, 2006. **368**(9530): p. 100-101.

20. Rossignol, J.F., et al., *Effect of nitazoxanide for treatment of severe rotavirus diarrhoea: randomised double-blind placebo-controlled trial.* Lancet, 2006. **368**(9530): p. 124-129.

21. Rossignol, J.F. and Y.M. El-Gohary, *Nitazoxanide in the treatment of viral gastroenteritis: a randomized double-blind placebo-controlled clinical trial.* Alimentary Pharmacology & Therapeutics, 2006. **24**(10): p. 1423-1430.

22. Callan, R.J., L.V. Ashton, and L.S. Goehring, *Nitazoxanide and tizoxanide inhibit EHV-1 and influenza type a virus replication in vitro.* Journal of Veterinary Internal Medicine, 2008. **22**(3): p. 738-738.

23. Muller, J., et al., *In vitro effects of thiazolides on Giardia lamblia WB clone C6 cultured axenically and in coculture with Caco2 cells.* Antimicrobial Agents and Chemotherapy, 2006. **50**(1): p. 162-170.

24. Hoffman, P.S., et al., *Antiparasitic drug nitazoxanide inhibits the pyruvate oxidoreductases of Helicobacter pylori, selected anaerobic bacteria and parasites, and Campylobacter jejuni.* Antimicrobial Agents and Chemotherapy, 2007. **51**(3): p. 868-876.

25. Korba, B., et al., *Potential Role for Nitazoxanide in Combination with Stat-C Agents for the Inhibition of Hcv Replication without the Development of Resistance.* Hepatology, 2008. **48**(4): p. 356a-356a.

26. Korba, B.E., et al., *Nitazoxanide, tizoxanide and other thiazolides are potent inhibitors of hepatitis B virus and hepatitis C virus replication.* Antiviral Res, 2008. **77**(1): p. 56-63.

27. Muller, J., et al., *Neospora caninum: Functional inhibition of protein disulfide isomerase by the broad-spectrum anti-parasitic drug nitazoxanide and other thiazolides.* Experimental Parasitology, 2008. **118**(1): p. 80-88.

28. Muller, J., et al., *Characterization of Giardia lamblia WB C6 clones resistant to nitazoxanide and to metronidazole.* J. Antimicrob. Chemother., 2007. **60**(2): p. 280-287.

29. Korba, B.E., et al., *Studies of the potential for resistance to nitazoxanide or tizoxanide.* Journal of Hepatology, 2008. **48**: p. S11-S11.

30. Muller, M., *Reductive activation of nitroimidazoles in anaerobic microorganisms.* Biochem Pharmacol, 1986. **35**(1): p. 37-41.

31. Townson, S.M., et al., *Resistance to the nitroheterocyclic drugs.* Acta Trop, 1994. **56**(2-3): p. 173-94.

32. Goldman, P., et al., *Comparing the reduction of nitroimidazoles in bacteria and mammalian tissues and relating it to biological activity.* Biochem Pharmacol, 1986. **35**(1): p. 43-51.

33. Dan, M., A.L. Wang, and C.C. Wang, *Inhibition of pyruvate-ferredoxin oxidoreductase gene expression in Giardia lamblia by a virus-mediated hammerhead ribozyme.* Mol Microbiol, 2000. **36**(2): p. 447-56.

34. Townson, S.M., J.A. Upcroft, and P. Upcroft, *Characterisation and purification of pyruvate:ferredoxin oxidoreductase from Giardia duodenalis.* Mol Biochem Parasitol, 1996. **79**(2): p. 183-93.

35. Leitsch, D., et al., *Trichomonas vaginalis: metronidazole and other nitroimidazole drugs are reduced by the flavin enzyme thioredoxin reductase and disrupt the cellular redox system. Implications for nitroimidazole toxicity and resistance.* Mol Microbiol, 2009. **72**(2): p. 518-36.

36. Adam, R.D., *Biology of Giardia lamblia.* Clin Microbiol Rev, 2001. **14**(3): p. 447-75.

37. Exner, M. and V. Gornik, *Parasitic zoonoses transmitted by drinking water. Giardiasis and cryptosporidiosis.* Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, 2004. **47**(7): p. 698-704.

38. Angarano, G., et al., *Giardiasis in HIV: a possible role in patients with severe immune deficiency.* Eur J Epidemiol, 1997. **13**(4): p. 485-7.

39. Feitosa, G., et al., *High prevalence of giardiasis and stronglyloidiasis among HIV-infected patients in Bahia, Brazil.* Braz J Infect Dis, 2001. **5**(6): p. 339-44.

40. Dobell, C., *The Discovery of the Intestinal Protozoa of Man.* Proc R Soc Med, 1920. **13**: p. 1-15.

41. Levine, N.D., et al., *A newly revised classification of the protozoa.* J Protozool, 1980. **27**(1): p. 37-58.

42. Leipe, D.D., et al., *Small subunit ribosomal RNA+ of Hexamita inflata and the quest for the first branch in the eukaryotic tree.* Mol Biochem Parasitol, 1993. **59**(1): p. 41-8.

43. Hashimoto, T., et al., *Phylogenetic place of mitochondrion-lacking protozoan, Giardia lamblia, inferred from amino acid sequences of elongation factor 2.* Mol Biol Evol, 1995. **12**(5): p. 782-93.

44. Hilario, E. and J.P. Gogarten, *The prokaryote-to-eukaryote transition reflected in the evolution of the V/F/A-ATPase catalytic and proteolipid subunits.* J Mol Evol, 1998. **46**(6): p. 703-15.

45. Roger, A.J., et al., *A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria.* Proc Natl Acad Sci U S A, 1998. **95**(1): p. 229-34.

46. Regoes, A., et al., *Protein import, replication, and inheritance of a vestigial mitochondrion.* J Biol Chem, 2005. **280**(34): p. 30557-63.

47. Logsdon, J.M., Jr., *Evolutionary genetics: sex happens in Giardia.* Curr Biol, 2008. **18**(2): p. R66-8.

48. Cooper, M.A., et al., *Population genetics provides evidence for recombination in Giardia.* Curr Biol, 2007. **17**(22): p. 1984-8.

49. Feely, D.E., J.V. Schollmeyer, and S.L. Erlandsen, *Giardia Spp - Distribution of Contractile Proteins in the Attachment Organelle.* Experimental Parasitology, 1982. **53**(1): p. 145-154.

50. Aggarwal, A. and T.E. Nash, *Characterization of a 33-Kilodalton Structural Protein of Giardia-Lamblia and Localization to the Ventral Disk.* Infection and Immunity, 1989. **57**(4): p. 1305-1310.

51. Soltys, B.J. and R.S. Gupta, *Immunoelectron Microscopy of Giardia-Lamblia Cytoskeleton Using Antibody to Acetylated Alpha-Tubulin.* Journal of Eukaryotic Microbiology, 1994. **41**(6): p. 625-632.

52. Campanati, L., et al., *Tubulin diversity in trophozoites of Giardia lamblia.* Histochemistry and Cell Biology, 2003. **119**(4): p. 323-331.

53. Elmendorf, H.G., S.C. Dawson, and M. McCaffery, *The cytoskeleton of Giardia lamblia.* International Journal for Parasitology, 2003. **33**(1): p. 3-28.

54. Feinstone, S.M., et al., *Non-A, non-B hepatitis in chimpanzees and marmosets.* J Infect Dis, 1981. **144**(6): p. 588-98.

55. Blanchard, E., et al., *Hepatitis C virus entry depends on clathrin-mediated endocytosis.* J Virol, 2006. **80**(14): p. 6964-72.

56. Otto, G.A. and J.D. Puglisi, *The pathway of HCV IRES-mediated translation initiation.* Cell, 2004. **119**(3): p. 369-80.

57. Branch, A.D., et al., *The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others.* Semin Liver Dis, 2005. **25**(1): p. 105-17.

58. Vassilaki, N. and P. Mavromara, *The HCV ARFP/F/core+1 protein: production and functional analysis of an unconventional viral product.* IUBMB Life, 2009. **61**(7): p. 739-52.

59. Mast, E.E., M.J. Alter, and H.S. Margolis, *Strategies to prevent and control hepatitis B and C virus infections: a global perspective.* Vaccine, 1999. **17**(13-14): p. 1730-3.

60. *Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium.* J Viral Hepat, 1999. **6**(1): p. 35-47.

61. Hadziyannis, S.J., et al., *Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose.* Ann Intern Med, 2004. **140**(5): p. 346-55.

62. Ascione, A., et al., *Peginterferon Alpha-2a Plus Ribavirin is More Effective than Peginterferon Alpha-2b Plus Ribavirin for Treating Chronic Hepatitis C Virus Infection.* Gastroenterology, 2009.

63. Sen, G.C., *Viruses and interferons.* Annu Rev Microbiol, 2001. **55**: p. 255-81.

64. Di Bisceglie, A.M. and J.H. Hoofnagle, *Optimal therapy of hepatitis C.* Hepatology, 2002. **36**(5 Suppl 1): p. S121-7.

65. Lau, J.Y., et al., *Mechanism of action of ribavirin in the combination treatment of chronic HCV infection.* Hepatology, 2002. **35**(5): p. 1002-9.

66. Maag, D., et al., *Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin.* J Biol Chem, 2001. **276**(49): p. 46094-8.

67. Contreras, A.M., et al., *Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system.* J Virol, 2002. **76**(17): p. 8505-17.

68. Crotty, S., C.E. Cameron, and R. Andino, *RNA virus error catastrophe: direct molecular test by using ribavirin.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6895-900.

69. Crotty, S., et al., *The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen.* Nat Med, 2000. **6**(12): p. 1375-9.

70. Meier, V. and G. Ramadori, *Hepatitis C virus virology and new treatment targets.* Expert Rev Anti Infect Ther, 2009. **7**(3): p. 329-50.

71. Lin, C., A.D. Kwong, and R.B. Perni, *Discovery and development of VX-950, a novel, covalent, and reversible inhibitor of hepatitis C virus NS3.4A serine protease.* Infect Disord Drug Targets, 2006. **6**(1): p. 3-16.

72. Berman, K. and P.Y. Kwo, *Boceprevir, an NS3 protease inhibitor of HCV.* Clin Liver Dis, 2009. **13**(3): p. 429-39.

73. Lohmann, V., et al., *Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.* Science, 1999. **285**(5424): p. 110-3.

74. Gry, M., et al., *Correlations between RNA and protein expression profiles in 23 human cell lines.* BMC Genomics, 2009. **10**: p. 365.

75. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.

76. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-63.

77. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.

78. Watson, J.D., *The human genome revealed.* Genome Res, 2001. **11**(11): p. 1803-4.

79. Fischer, B., et al., *NovoHMM: a hidden Markov model for de novo peptide sequencing.* Anal Chem, 2005. **77**(22): p. 7265-73.

80. Dancik, V., et al., *De novo peptide sequencing via tandem mass spectrometry.* J Comput Biol, 1999. **6**(3-4): p. 327-42.

81. Chen, T., et al., *A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry.* J Comput Biol, 2001. **8**(3): p. 325-37.

82. Thomson, J.M., et al., *A custom microarray platform for analysis of microRNA gene expression.* Nat Methods, 2004. **1**(1): p. 47-53.

83. Shendure, J., *The beginning of the end for microarrays?* Nat Methods, 2008. **5**(7): p. 585-7.

84. Hartinger, J., et al., *16-BAC/SDS-PAGE: a two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins.* Anal Biochem, 1996. **240**(1): p. 126-33.

85. Campostrini, N., et al., *Spot overlapping in two-dimensional maps: a serious problem ignored for much too long.* Proteomics, 2005. **5**(9): p. 2385-95.

86. Pietrogrande, M.C., et al., *Spot overlapping in two-dimensional polyacrylamide gel electrophoresis maps: relevance to proteomics.* Electrophoresis, 2003. **24**(1-2): p. 217-24.

87. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

88. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.

89. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.

90. *The Giardia lamblia Genome Database*. 2006.

91. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-3567.

92. W.H.O., *Intestinal parasitic infections*, in *Global health situation and projections-estimates*. 1992, World Health Organization: Geneva. p. 46.

93. Fraser, D., et al., *Giardia lamblia carriage in Israeli Bedouin infants: Risk factors and consequences.* Clinical Infectious Diseases, 2000. **30**(3): p. 419-424.

94. Sisson, G., et al., *Enzymes associated with reductive activation and action of nitazoxanide, nitrofurans, and metronidazole in Helicobacter pylori.* Antimicrobial Agents and Chemotherapy, 2002. **46**(7): p. 2116-2123.

95. Drummelsmith, J., et al., *Proteome mapping of the protozoan parasite Leishmania and application to the study of drug targets and resistance mechanisms.* Mol Cell Proteomics, 2003. **2**(3): p. 146-55.

96. Di Michele, M., et al., *A proteomic approach to paclitaxel chemoresistance in ovarian cancer cell lines.* Biochim Biophys Acta, 2009. **1794**(2): p. 225-36.

97. Lis, M. and L.A. Bobek, *Proteomic and metabolic characterization of a Candida albicans mutant resistant to the antimicrobial peptide MUC7 12-mer.* FEMS Immunol Med Microbiol, 2008. **54**(1): p. 80-91.

98. Fernandez-Reyes, M., et al., *The cost of resistance to colistin in Acinetobacter baumannii: a proteomic perspective.* Proteomics, 2009. **9**(6): p. 1632-45.

99. Biller, L., et al., *Comparison of two genetically related Entamoeba histolytica cell lines derived from the same isolate with different pathogenic properties.* Proteomics, 2009. **9**(17): p. 4107-20.

100. Albert, T.J., et al., *Mutation discovery in bacterial genomes: metronidazole resistance in Helicobacter pylori.* Nat Methods, 2005. **2**(12): p. 951-3.

101. Davis, B.M. and M.K. Waldor, *High-throughput sequencing reveals suppressors of Vibrio cholerae rpoE mutations: one fewer porin is enough.* Nucleic Acids Res, 2009. **37**(17): p. 5757-67.

102. Chepelev, I., et al., *Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq.* Nucleic Acids Res, 2009. **37**(16): p. e106.

103. Keister, D.B., *Axenic Culture of Giardia-Lamblia in Tyi-S-33 Medium Supplemented with Bile.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1983. **77**(4): p. 487-488.

104. Morrison, H.G., et al., *Genomic minimalism in the early diverging intestinal parasite Giardia lamblia.* Science, 2007. **317**(5846): p. 1921-6.

105. Wang, Z., G.J. Vora, and D.A. Stenger, *Detection and genotyping of Entamoeba histolytica, Entamoeba dispar, Giardia lamblia, and Cryptosporidium parvum by oligonucleotide microarray.* J Clin Microbiol, 2004. **42**(7): p. 3262-71.

106. Muller, J., et al., *Identification of differentially expressed genes in a Giardia lamblia WBC6 clone resistant to nitazoxanide and metronidazole.* Journal of Antimicrobial Chemotherapy, 2008. **62**(1): p. 72-82.

107. Roxstrom-Lindquist, K., et al., *Giardia lamblia-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells.* Infect Immun, 2005. **73**(12): p. 8204-8.

108. Steuart, R.F., *Proteomic analysis of Giardia: Studies from the pre- and post-genomic era.* Exp Parasitol, 2009.

109. Xia, D., et al., *The proteome of Toxoplasma gondii: integration with the genome provides novel insights into gene expression and annotation.* Genome Biology, 2008. **9**(7): p. -.

110. Keister, D.B., *Axenic culture of Giardia lamblia in TYI-S-33 medium supplemented with bile.* Trans R Soc Trop Med Hyg, 1983. **77**(4): p. 487-8.

111. Team, R.D.C., *R: A language and environment for statistical computing.* 2008, Vienna, Austria.

112. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.

113. Smyth, G.K., J. Michaud, and H.S. Scott, *Use of within-array replicate spots for assessing differential expression in microarray experiments.* Bioinformatics, 2005. **21**(9): p. 2067-75.

114. Smyth, G.K. and T. Speed, *Normalization of cDNA microarray data.* Methods, 2003. **31**(4): p. 265-73.

115. Wettenhall, J.M. and G.K. Smyth, *limmaGUI: a graphical user interface for linear modeling of microarray data.* Bioinformatics, 2004. **20**(18): p. 3705-6.

116. Kim, J., et al., *Comparative proteomic analysis of trophozoites versus cysts of Giardia lamblia.* Parasitol Res, 2009. **104**(2): p. 475-9.

117. Bjellqvist, B., et al., *The Focusing Positions of Polypeptides in Immobilized Ph Gradients Can Be Predicted from Their Amino-Acid-Sequences.* Electrophoresis, 1993. **14**(10): p. 1023-1031.

118. Lechtreck, K.F., J. Rostmann, and A. Grunow, *Analysis of Chlamydomonas SF-assemblin by GFP tagging and expression of antisense constructs.* Journal of Cell Science, 2002. **115**(7): p. 1511-1522.

119. Lechtreck, K.F. and C.D. Silflow, *SF-assemblin in Chlamydomonas: Sequence conservation and localization during the cell cycle.* Cell Motility and the Cytoskeleton, 1997. **36**(2): p. 190-201.

120. Solari, A.J., et al., *A unique mechanism of nuclear division in Giardia lamblia involves components of the ventral disk and the nuclear envelope.* Biocell, 2003. **27**(3): p. 329-346.

121. Benchimol, M., *Participation of the Adhesive Disc during Karyokinesis in Giardia lamblia.* Biology of the Cell, 2004. **96**(4): p. 291-301.

122. Chakraborty, P., et al., *Nucleoporin levels regulate cell cycle progression and phase-specific gene expression.* Dev Cell, 2008. **15**(5): p. 657-67.

123. Xu, S. and M.A. Powers, *Nuclear pore proteins and cancer.* Semin Cell Dev Biol, 2009. **20**(5): p. 620-30.

124. Sagolla, M.S., et al., *Three-dimensional analysis of mitosis and cytokinesis in the binucleate parasite Giardia intestinalis.* J Cell Sci, 2006. **119**(Pt 23): p. 4889-900.

125. Svard, S.G., et al., *Differentiation-associated surface antigen variation in the ancient eukaryote Giardia lamblia.* Mol Microbiol, 1998. **30**(5): p. 979-89.

126. Bader, N. and T. Grune, *Protein oxidation and proteolysis.* Biol Chem, 2006. **387**(10-11): p. 1351-5.

127. Shringarpure, R., et al., *Ubiquitin conjugation is not required for the degradation of oxidized proteins by proteasome.* J Biol Chem, 2003. **278**(1): p. 311-8.

128. Nash, T.E., *Antigenic variation in Giardia lamblia and the host's immune response.* Philos Trans R Soc Lond B Biol Sci, 1997. **352**(1359): p. 1369-75.

129. Rossignol, J.F., A. Elfert, and E.B. Keeffe, *Evaluation of a 4 week lead-in phase with nitazoxanide prior to nitazoxanide plus peginterferon in treating chronic hepatitis C.* Journal of Hepatology, 2008. **48**: p. S312-S312.

130. Rossignol, J.F., et al., *Improved virologic response in chronic hepatitis C genotype 4 treated with nitazoxanide, peginterferon, and ribavirin.* Gastroenterology, 2009. **136**(3): p. 856-62.

131. Elazor, M., et al., *Nitazoxanide (Ntz) Is an Inducer Eif2a and Pkr Phosphorylation.* Hepatology, 2008. **48**(4): p. 1151a-1151a.

132. Elazar, M., et al., *The Anti-Hepatitis C Agent Nitazoxanide Induces Phosphorylation of Elongation Initiation Factor 2alpha Via Protein Kinase Activated by Double-Stranded RNA Activation.* Gastroenterology, 2009.

133. Clemens, M.J. and A. Elia, *The double-stranded RNA-dependent protein kinase PKR: structure and function.* J Interferon Cytokine Res, 1997. **17**(9): p. 503-24.

134. Kaufman, R.J., *Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls.* Genes Dev, 1999. **13**(10): p. 1211-33.

135. Mori, K., *Tripartite management of unfolded proteins in the endoplasmic reticulum.* Cell, 2000. **101**(5): p. 451-4.

136. Schroder, M. and R.J. Kaufman, *The mammalian unfolded protein response.* Annu Rev Biochem, 2005. **74**: p. 739-89.

137. Urano, F., et al., *Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1.* Science, 2000. **287**(5453): p. 664-6.

138. Chen, X., J. Shen, and R. Prywes, *The luminal domain of ATF6 senses endoplasmic reticulum (ER) stress and causes translocation of ATF6 from the ER to the Golgi.* J Biol Chem, 2002. **277**(15): p. 13045-52.

139. Okada, T., et al., *A serine protease inhibitor prevents endoplasmic reticulum stress-induced cleavage but not transport of the membrane-bound transcription factor ATF6.* J Biol Chem, 2003. **278**(33): p. 31024-32.

140. Kokame, K., H. Kato, and T. Miyata, *Identification of ERSE-II, a new cis-acting element responsible for the ATF6-dependent mammalian unfolded protein response.* J Biol Chem, 2001. **276**(12): p. 9199-205.

141. Yoshida, H., et al., *Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional*

induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. J Biol Chem, 1998. **273**(50): p. 33741-9.

142. Li, S.S., et al., *Hepatitis C virus NS4B induces unfolded protein response and endoplasmic reticulum overload response-dependent NF-kappa B activation.* Virology, 2009. **391**(2): p. 257-264.

143. Chan, S.W. and P.A. Egan, *Hepatitis C virus envelope proteins regulate CHOP via induction of the unfolded protein response.* Faseb Journal, 2005. **19**(9): p. 1510-+.

144. Benali-Furet, N.L., et al., *Hepatitis C virus core triggers apoptosis in liver cells by inducing ER stress and ER calcium depletion.* Oncogene, 2005. **24**(31): p. 4921-4933.

145. Sir, D., et al., *Induction of incomplete autophagic response by hepatitis C virus via the unfolded protein response.* Hepatology, 2008. **48**(4): p. 1054-1061.

146. Gale, M.J., et al., *Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein.* Virology, 1997. **230**(2): p. 217-227.

147. Taylor, D.R., et al., *Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein.* Science, 1999. **285**(5424): p. 107-110.

148. Tardif, K.D., et al., *Hepatitis C virus suppresses the IRE1-XBP1 pathway of the unfolded protein response.* Journal of Biological Chemistry, 2004. **279**(17): p. 17158-17164.

149. Elazar, M., et al., *Amphipathic helix-dependent localization of NS5A mediates hepatitis C virus RNA replication.* J Virol, 2003. **77**(10): p. 6055-61.

150. Tan, S.L., et al., *Hepatitis C therapeutics: current status and emerging strategies.* Nat Rev Drug Discov, 2002. **1**(11): p. 867-81.

151. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.

152. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

153. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

154. Jensen, L.J., et al., *STRING 8--a global view on proteins and their functional interactions in 630 organisms.* Nucleic Acids Res, 2009. **37**(Database issue): p. D412-6.

155. Tardif, K.D., K. Mori, and A. Siddiqui, *Hepatitis C virus subgenomic replicons induce endoplasmic reticulum stress activating an intracellular signaling pathway.* J Virol, 2002. **76**(15): p. 7453-9.

156. Lee, A.H., N.N. Iwakoshi, and L.H. Glimcher, *XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response.* Mol Cell Biol, 2003. **23**(21): p. 7448-59.

157. Fang, C., et al., *Proteome analysis of human liver carcinoma Huh7 cells harboring hepatitis C virus subgenomic replicon.* Proteomics, 2006. **6**(2): p. 519-27.

158. Lai, C.K., et al., *Association of hepatitis C virus replication complexes with microtubules and actin filaments is dependent on the interaction of NS3 and NS5A.* J Virol, 2008. **82**(17): p. 8838-48.

159. Wolk, B., et al., *A dynamic view of hepatitis C virus replication complexes.* J Virol, 2008. **82**(21): p. 10519-31.

160. Liu, Z., et al., *Critical role of cyclophilin A and its prolyl-peptidyl isomerase activity in the structure and function of the hepatitis C virus replication complex.* J Virol, 2009. **83**(13): p. 6554-65.

161. Hara, H., et al., *Involvement of creatine kinase B in hepatitis C virus genome replication through interaction with the viral NS4A protein.* J Virol, 2009. **83**(10): p. 5137-47.

162. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.

163. Green, N.M., *Avidin. 1. The Use of (14-C)Biotin for Kinetic Studies and for Assay.* Biochem J, 1963. **89**: p. 585-91.

164. Bayer, E.A. and M. Wilchek, *The use of the avidin-biotin complex as a tool in molecular biology.* Methods of biochemical analysis, 1980. **26**: p. 1-45.

165. Muller, J., et al., *A novel Giardia lamblia nitroreductase, GlNR1, interacts with nitazoxanide and other thiazolides.* Antimicrobial Agents and Chemotherapy, 2007. **51**(6): p. 1979-1986.

166. Muller, J., et al., *Thiazolides inhibit growth and induce glutathione-S-transferase Pi (GSTP1)-dependent cell death in human colon cancer cells.* Int J Cancer, 2008. **123**(8): p. 1797-806.

167. *Affinity Chromatography: Principles and Methods*. Principles and Methods. 2007, Uppsala: GE Healthcare Bio-Sciences AB. 159.

168. Laemmli, U.K., *Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4.* Nature, 1970. **227**(5259): p. 680-&.

169. Bantscheff, M., et al., *Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors.* Nat Biotechnol, 2007. **25**(9): p. 1035-44.

170. von Rechenberg, M., et al., *Ampicillin/penicillin-binding protein interactions as a model drug-target system to optimize affinity pull-down and mass spectrometric strategies for target and pathway identification.* Proteomics, 2005. **5**(7): p. 1764-73.

171. Sato, S., et al., *Polyproline-rod approach to isolating protein targets of bioactive small molecules: isolation of a new target of indomethacin.* J Am Chem Soc, 2007. **129**(4): p. 873-80.

172. Chase, H.A., *Purification of proteins by adsorption chromatography in expanded beds.* Trends Biotechnol, 1994. **12**(8): p. 296-303.

173. Harris, E.L.V. and S. Angal, *Protein purification methods : a practical approach.* Practical approach series. 1989, Oxford: Irl. xvi, 317.

174. Sodhi, R. and Y.S. Rajput, *Method for dialysis of samples in microliter volumes.* Anal Biochem, 2003. **315**(1): p. 141-2.

175. Overall, C.M., *A microtechnique for dialysis of small volume solutions with quantitative recoveries.* Anal Biochem, 1987. **165**(1): p. 208-14.

176. Horowitz, P.M. and L.D. Barnes, *A simple, inexpensive, and precise microcell for the exchange dialysis and equilibrium dialysis of small samples.* Anal Biochem, 1983. **128**(2): p. 478-80.

177. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angew Chem Int Ed Engl, 2001. **40**(11): p. 2004-2021.

178. Faig, M., et al., *Structures of recombinant human and mouse NAD(P)H : quinone oxidoreductases: Species comparison and structural changes with substrate binding and release.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(7): p. 3177-3182.

179. Moran, J.L., D. Siegel, and D. Ross, *A potential mechanism underlying the increased susceptibility of individuals with a polymorphism in NAD(P)H : quinone oxidoreductase 1 (NQO1) to benzene toxicity.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(14): p. 8150-8155.

180. Asher, G., et al., *Regulation of p53 stability and p53-dependent apoptosis by NADH quinone oxidoreductase 1.* PNAS, 2001. **98**(3): p. 1188-1193.

181. Diehn, M., et al., *SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data.* Nucleic Acids Research, 2003. **31**(1): p. 219-223.

182. Ross, D., et al., *NAD(P)H:quinone oxidoreductase 1 (NQO1): chemoprotection, bioactivation, gene regulation and genetic polymorphisms.* Chemico-Biological Interactions, 2000. **129**(1-2): p. 77.

183. Colucci, M.A., et al., *Synthesis and evaluation of 3-aryloxymethyl-1,2-dimethylindole-4,7-diones as mechanism-based inhibitors of NAD(P)H : Quinone oxidoreductase 1 (NQO1) activity.* Journal of Medicinal Chemistry, 2007. **50**(23): p. 5780-5789.

184. Jaiswal, A.K., *Human Nad(P)H - Quinone Oxidoreductase (Nqo1) Gene Structure and Induction by Dioxin.* Biochemistry, 1991. **30**(44): p. 10647-10653.

185. Ernster, L., W.E. Ronald, and E.P. Maynard, *DT diaphorase*, in *Methods in Enzymology*. 1967, Academic Press. p. 309-317.

186. Lind, C., et al., *Dt-Diaphorase - Purification, Properties, and Function.* Methods in Enzymology, 1990. **186**: p. 287-301.

187. Laskowski, R.A., et al., *Procheck - a Program to Check the Stereochemical Quality of Protein Structures.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

188. Eisenberg, D., R. Luthy, and J.U. Bowie, *VERIFY3D: Assessment of protein models with three-dimensional profiles.* Macromolecular Crystallography, Pt B, 1997. **277**: p. 396-404.

189. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Journal of Molecular Modeling, 2001. **7**(8): p. 306-317.

190. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.

191. Tipton, K.F., *Enzyme Assays: A Practical Approach*, in *The Practical Approach Series*, R.a.D. Eisenthal, M. J., Editor. 1992, Oxford University Press: Oxford. p. 1 - 58.

192. Tsvetkov, P., et al., *Inhibition of NAD(P)H:quinone oxidoreductase 1 activity and induction of p53 degradation by the natural phenolic*

compound curcumin. Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5535-40.

193. Gliszczynska-Swiglo, A., et al., *The role of quinone reductase (NQO1) and quinone chemistry in quercetin cytotoxicity.* Toxicology in Vitro, 2003. **17**(4): p. 423-431.

194. Chen, S., et al., *Molecular Characterization of Binding of Substrates and Inhibitors to DT-Diaphorase: Combined Approach Involving Site-Directed Mutagenesis, Inhibitor-Binding Analysis, and Computer Modeling.* Mol Pharmacol, 1999. **56**(2): p. 272-278.

195. Lan, K.H., et al., *HCVNS5A interacts with p53 and inhibits p53-mediated apoptosis.* Oncogene, 2002. **21**(31): p. 4801-4811.

196. Muller, J., et al., *Thiazolides inhibit growth and induce glutathione-S-transferase Pi (GSTP1)-dependent cell death in human colon cancer cells.* International Journal of Cancer, 2008. **123**(8): p. 1797-1806.

197. Zhang, F., et al., *Ribosomal stress couples the unfolded protein response to p53-dependent cell cycle arrest.* Journal of Biological Chemistry, 2006. **281**(40): p. 30036-30045.

198. Tardif, K.D., G. Waris, and A. Siddiqui, *Hepatitis C virus, ER stress, and oxidative stress.* Trends in Microbiology, 2005. **13**(4): p. 159-163.

199. Abboud, P., et al., *Successful treatment of metronidazole- and albendazole-resistant giardiasis with nitazoxanide in a patient with acquired immunodeficiency syndrome.* Clinical Infectious Diseases, 2001. **32**(12): p. 1792-1794.

200. Theodos, C.M., et al., *Efficacy of nitazoxanide against Cryptosporidium parvum in cell culture and in animal models.* Antimicrobial Agents and Chemotherapy, 1998. **42**(8): p. 1959-1965.

201. Rossignol, J.F., et al., *Thiazolides, a new class of anti-influenza molecules targeting viral hemagglutinin at post-translational level.* J Biol Chem, 2009.

202. *The Database for Annotation, Visualization and Integrated Discovery.* 2009 [cited 2009 November]; Private beta release of new version of DAVID]. Available from: http://david.abcc.ncifcrf.gov:8080/.

203. Asher, G., et al., *The crystal structure of NAD(P)H quinone oxidoreductase 1 in complex with its potent inhibitor dicoumarol.* Biochemistry, 2006. **45**(20): p. 6372-8.

204. Jaffar, M., et al., *3-substituted-5-aziridinyl-1-methylindole-4,7-diones as NQO1-directed antitumour agents: mechanism of activation and cytotoxicity in vitro.* Biochem Pharmacol, 2003. **66**(7): p. 1199-206.

205. Pietschmann, T., et al., *Persistent and transient replication of full-length hepatitis C virus genomes in cell culture.* J Virol, 2002. **76**(8): p. 4008-21.

206. Florens, L., et al., *A proteomic view of the Plasmodium falciparum life cycle.* Nature, 2002. **419**(6906): p. 520-6.

207. Sanderson, S.J., et al., *Determining the protein repertoire of Cryptosporidium parvum sporozoites.* Proteomics, 2008. **8**(7): p. 1398-414.

208. Yates, J.R., 3rd, et al., *Direct analysis of protein mixtures by tandem mass spectrometry.* J Protein Chem, 1997. **16**(5): p. 495-7.

209. Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.

210. Raggiaschi, R., S. Gotta, and G.C. Terstappen, *Phosphoproteome analysis.* Biosci Rep, 2005. **25**(1-2): p. 33-44.

211. Ramachandran, N., et al., *Next-generation high-density self-assembling functional protein arrays.* Nat Methods, 2008. **5**(6): p. 535-8.

212. He, M., et al., *Printing protein arrays from DNA arrays.* Nat Methods, 2008. **5**(2): p. 175-7.

213. Kitteringham, N.R., et al., *Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics.* J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(13): p. 1229-39.

214. Cross, G.H., et al., *A new quantitative optical biosensor for protein characterisation.* Biosens Bioelectron, 2003. **19**(4): p. 383-90.

215. Bertucci, C. and S. Cimitan, *Rapid screening of small ligand affinity to human serum albumin by an optical biosensor.* J Pharm Biomed Anal, 2003. **32**(4-5): p. 707-14.

Appendix A

"TOPTABLE" RANKED LIST OF DIFFERENTIALLY EXPRESSED GENES IN NTZ RESISTANT *GIARDIA LAMBLIA*

locus	[Product Description]	logFC
113797	VSP with INR	-3.3826208
4078	Hypothetical protein	1.9680186
16501	VSP with INR	1.8392107
111874	VSP	-2.1392015
99493	hypothetical protein	-2.6984578
114122	VSP	-1.6893063
41212	Hypothetical protein	1.4018316
5435	Cyst wall protein 2	-1.4520654
137612	VSP	1.3231786
114672	VSP	-1.5510569
101811	Hypothetical protein	-1.5329691
114065	VSP	-1.788614
40591	VSP	-1.2666862
40621	VSP	-1.3834347
17249	Coiled-coil protein	1.4868398
36493	VSP	-1.6011802
98799	Hypothetical protein	1.9836519
10429	Wos2 protein	2.2770209
32450	hypothetical protein	-1.2008322
32426	Hypothetical protein	1.4082055
111873	VSP	-1.2664834
27310	Stress-induced-phosphoprotein 1	1.4193631
7593	Hypothetical protein	1.2907629
118532	hypothetical protein	-1.1511618
101326	Hypothetical protein	1.2746888
118156	hypothetical protein	-1.4314666
13520	VSP	-1.1932273
12143	hypothetical protein	-1.0693958
97733	Hypothetical protein	-1.618966
3218	hypothetical protein	-1.1842615
7260	Aldose reductase	1.1169795
4812	Beta-giardin	2.1184591
12139	Protein 21.1	2.2388346
9861	Hypothetical protein	1.1166114
8722	Myb 1-like protein	-1.1381169
34030	hypothetical protein	-1.2260634
115200	Hypothetical protein	-1.002655
137672	High cysteine membrane protein VSP-like	-1.2245806
27925	Protein 21.1	2.448497
137740	VSP	1.4999011
13272	Hypothetical protein	-1.5441317

112103	Arginine deiminase	1.6544307
10193	WD-repeat membrane protein, putative	-1.2746311
41476	VSP	-1.3313304
17043	Glyceraldehyde 3-phosphate dehydrogenase	-1.0129174
28748	Hypothetical protein	1.0772775
106300	hypothetical protein	-1.0371954
26971	hypothetical protein	-1.1695203
16745	Axoneme-associated protein GASP-180	1.253151
8174	Protein 21.1	1.4127078
5188	Protein 21.1	1.6739594
18877	hypothetical protein	-1.8815889
115598	hypothetical protein	-1.2910638
33004	Hypothetical protein	-1.4381026
1206	hypothetical protein	1.1189112
38862	hypothetical protein	-1.0444742
15026	Src-associated protein-like protein	1.1591176
92941	hypothetical protein	1.1637746
137713	Hypothetical protein	-2.1353091
24590	Protein 21.1	1.0980898
29774	Hypothetical protein	-1.2976241
101498	Variant-specific surface protein VSP4A1 precursor	-1.4840273
19098	hypothetical protein	-3.4584075
97820	VSP, putative	-2.2375817
8173	Glycerol kinase	-1.4510159
32711	hypothetical protein	1.067822
13783	Hypothetical protein	-1.0429467
16125	Glycerol-3-phosphate dehydrogenase	-1.2392267
102012	Hypothetical protein	-1.0581393
8726	Actin related protein	1.6191699
38553	Hypothetical protein	1.2161621
16906	Phosphatidate cytidylyltransferase	1.0845438
14043	VSP	-1.343418
5638	Cyst wall protein 1	-1.1015245
134773	hypothetical protein	-1.0466678
114653	VSP	-1.2179331
34196	VSP	-1.03784
17551	Protein 21.1	2.2383348
17570	Elongation factor 2	1.0974692
16648	Hypothetical protein	1.1729272
11865	Hypothetical protein	1.0868362
4410	SALP-1	1.30995
16283	Amino acid transporter family	-1.4905046

135831	VSP	-1.018952
13627	Hypothetical protein	1.0077234
19663	Hypothetical protein	-1.1426454
103983	Hypothetical protein	-1.6084021
11997	hypothetical protein	-1.0510821
21444	Spindle pole protein, putative	1.6790508
32265	Hypothetical protein	1.6550273
15455	E04F6.2 like protein	1.0677176
16532	Protein 21.1	2.3858302
137726	ABC transporter ABCA.1, putative	-1.1373928
14262	Hypothetical protein	1.4405229
117476	Hypothetical protein	1.0310408
23876	Hypothetical protein	1.1565427
34867	hypothetical protein	1.1266217
6334	Hypothetical protein	-1.1877477
137614	VSP AS12	1.1781061
17053	Protein 21.1	1.8847839
88369	Protein 21.1	1.2958509
13651	Hypothetical protein	1.1767041
10401	hypothetical protein	1.4399468
31692	Hypothetical protein	1.2881158
10255	Translation initiation factor eIF-4A, putative	1.107429
16412	Heat-shock protein, putative	1.0878148
113622	Protein 21.1	1.7593508
114623	Hypothetical protein	1.0264678
125386	hypothetical protein	-1.1708098
16867	AAA family ATPase	1.4665034
112445	Hypothetical protein	-1.0908036
98058	VSP	-1.5229418
4852	Hypothetical protein	1.3981064
31366	Hypothetical protein	-1.7030155
9515	Coiled-coil protein	1.3591797
13584	Hypothetical protein	1.1238364
10219	Protein 21.1	1.0802242
88765	Cytosolic HSP70	2.5871322
15499	Hypothetical protein	1.1508821
86648	Hypothetical protein	1.5631548
86484	Hypothetical protein	1.3748453
135918	VSP	-1.0393734
9978	Hypothetical protein	-1.1653574
10133	Hypothetical protein	-1.0019661
38906	Hypothetical protein	1.0061533

112048	VSP	1.4267342
5883	Hypothetical protein	1.556997
13329	Hypothetical protein	1.2893323
9046	Sugar transport family protein	-1.1510486
23492	Protein 21.1	1.3657967
1695	Rab11	1.0229949
23017	Hypothetical protein	1.516429
109417	hypothetical protein	-1.0348212
113954	VSP	-1.0442842
35011	Hypothetical protein	1.2213966
37538	Hypothetical protein	1.3461582
115985	hypothetical protein	-1.2471546
113450	VSP with INR	-2.7001563
3762	Protein 21.1	1.1465729
13323	Hypothetical protein	1.0192144
93011	Protein 21.1	1.3267276
97839	hypothetical protein	-1.2776117
17097	Protein 21.1	1.0327236
9720	Protein 21.1	1.1685414
10808	Hypothetical protein	1.0376391
93585	hypothetical protein	-1.094157
16844	Hypothetical protein	1.6260784
6637	hypothetical protein	-1.3347951
28811	hypothetical protein	-1.9846866
10238	Hypothetical protein	1.4219284
14628	Phosphatidylinositol-4-phosphate 5-kinase, putative	1.0022832
10524	Hypothetical protein	1.0858196
8044	Seven transmembrane protein 1	1.1686363
17046	Protein 21.1	2.0505091
14225	CXC-rich protein	-1.0050698
14940	Hypothetical protein	1.0329953
98126	High cysteine protein	-1.4017662
7439	Ser/Thr phosphatase 2A, 65kDa reg sub A	1.6941958
119466	hypothetical protein	-1.1417615
39024	hypothetical protein	-1.3802749
8377	Hypothetical protein	-2.4310123
4587	Hypothetical protein	1.0597201
10370	ATP/GTP binding protein, putative	1.0744608
15410	Ser/Thr protein kinase	1.2220712
16935	Hypothetical protein	1.4373532
14568	Ser/Thr protein phosphatase PP1-alpha 2 catalytic subunit	1.2900581
116967	Hypothetical protein	1.746827

2616	Hypothetical protein	1.1234678
114662	Hypothetical protein	1.0958169
3545	hypothetical protein	-1.1242322
20810	hypothetical protein	-2.004941
23790	Hypothetical protein	-1.0262794
36567	Hypothetical protein	1.5390717
13948	Hypothetical protein	-1.3783046
16376	ATP-dependent RNA helicase p47, putative	1.0479762
106224	hypothetical protein	-1.0196269
38761	Hypothetical protein	-1.1520153
103807	Protein 21.1	1.049963
24108	hypothetical protein	-1.4281621
102142	Hypothetical protein	-1.4407456
20170	hypothetical protein	-1.0494125
9787	Hypothetical protein	-1.0788858
32052	hypothetical protein	-1.0582137
24781	Hypothetical protein	-1.0170255
17060	Protein 21.1	1.8379919
24279	Hypothetical protein	1.05349
96818	Hypothetical protein	1.0959669
41401	VSP	-1.1236116
28302	hypothetical protein	-1.1938622
9062	Long chain fatty acid CoA ligase 5	1.0569044
112112	Hypothetical protein	1.5300598
16367	Hypothetical protein	1.2941586
3079	Hypothetical protein	-1.3326145
14859	Protein 21.1	2.0746506
6967	hypothetical protein	-1.169392
30947	hypothetical protein	-1.0787223
17585	Protein 21.1	1.3168111
16267	Hypothetical protein	1.1280021
40390	Protein 21.1	1.0964091
117118	hypothetical protein	-1.0488257
15048	ATP-dependent RNA helicase-like protein	1.5600422
95451	Hypothetical protein	-1.9826522
117908	hypothetical protein	-1.022446
86676	Delta giardin	1.2288563
17571	Trichohyalin	1.3361601
38009	Hypothetical protein	-1.2123463
28626	VSP	-1.1149377
39098	Hypothetical protein	-1.5422948
11309	High cysteine membrane protein Group 1	1.6655639

11165	Protein 21.1	1.0531813
103107	Hypothetical protein	-1.6750535
8855	P115, putative	1.0241137
7268	Protein 21.1	1.4136421
6242	Translationally controlled tumor protein-like protein	1.1911214
15823	1,4-alpha-glucan branching enzyme	1.2206884
14553	CCAAT-binding transcription factor subunit C	-1.2477901
41288	Mucin-like protein	1.001736
89117	Hypothetical protein	-1.0714499
26121	hypothetical protein	-1.0621025
11204	Plant adhesion molecule 1	-1.4158823
28829	hypothetical protein	1.1731792
40376	High cysteine non-variant cyst protein	-1.0214051
27896	hypothetical protein	1.1167383
16478	Hypothetical protein	1.3779455
38910	VSP	2.350619
5804	hypothetical protein	-1.950971
12229	Hypothetical protein	-1.7484415
32111	hypothetical protein	-1.7396331
31341	hypothetical protein	-1.8716566
114820	Hypothetical protein	-1.1366388
118283	Hypothetical protein	-1.3118835
10706	hypothetical protein	-1.0180786
5278	Hypothetical protein	-1.0865194
88181	Hypothetical protein	-1.2919539
36036	Hypothetical protein	-1.1481444
8364	Thymidine kinase	2.1832573
32051	Hypothetical protein	-1.7035587
3582	Hypothetical protein	1.2514338
34505	hypothetical protein	-1.7591143
96043	Hypothetical protein	-1.9836693
28348	hypothetical protein	-1.4336243
103818	Hypothetical protein	1.0735893
17230	Gamma giardin	2.2381868
27830	hypothetical protein	-1.0955445
102662	VSP	1.0911661
28761	Hypothetical protein	-1.3022835
12066	hypothetical protein	-1.8361274
33935	Hypothetical protein	-1.0825331
10171	hypothetical protein	-2.6249601
17120	CEGP1 protein	-1.1624368
4355	Nif3-related protein	-1.1613566

114674	Hypothetical protein	1.7982051
87941	hypothetical protein	-2.1591043
20386	hypothetical protein	-1.1488136
38412	Hypothetical protein	-1.9033777
2178	Hypothetical protein	-1.0011942
111809	Hypothetical protein	-1.0273758
92037	hypothetical protein	-2.5301086
93369	Hypothetical protein	-1.1380863
14438	hypothetical protein	-1.0177607
27760	hypothetical protein	-1.110506
32192	hypothetical protein	-3.2889784
13268	Hypothetical protein	-1.7255415
14971	Hypothetical protein	1.2278697
5302	hypothetical protein	-1.5575743
36798	Hypothetical protein	-1.1813553
16001	Glutamate-cysteine ligase	-1.5789414
102051	Amino acid transporter family	-2.0948855
114833	hypothetical protein	1.2520665
113021	Acetyl-CoA carboxylase/pyruvate carboxylase fusion protein, putative	-1.1265481
10361	P60 katanin	-1.4344822
10229	Hypothetical protein	-1.124833
92449	hypothetical protein	-1.0525987
37897	Hypothetical protein	-1.7525784
105663	hypothetical protein	-1.8173938
92990	Hypothetical protein	-4.1855426
94534	Hypothetical protein	1.0702163
12068	Hypothetical protein	-1.6567428
86401	hypothetical protein	-1.3783849
32354	hypothetical protein	-1.3289421
135991	Hypothetical protein	-2.3683503
37641	Hypothetical protein	-1.0067555
3273	hypothetical protein	-1.68979
13464	Splicesome-associated protein	-1.4899715
29599	hypothetical protein	-2.4587618
33978	Hypothetical protein	-1.0322838
38396	hypothetical protein	-1.2888904
24169	hypothetical protein	-1.1897364
39564	hypothetical protein	1.3632847
14821	HesB domain-containing protein	-1.2796581
31504	hypothetical protein	-3.4109045
8101	Hypothetical protein	-1.4910123
6602	Protein 21.1	-1.5423632

9238	hypothetical protein	-1.3456202
39487	hypothetical protein	-1.326162
38901	VSP	1.2140658
117147	hypothetical protein	-1.2377587
103164	H-SHIPPO 1	1.5361031
5513	Hypothetical protein	-1.393345
36178	hypothetical protein	-1.0159547
102184	Hypothetical protein	1.927578
31541	hypothetical protein	-1.1088954
34559	hypothetical protein	-1.3525332
10666	Histone acetyltransferase GCN5	-1.2064225
7053	Hypothetical protein	-2.0003323
21438	hypothetical protein	-1.7091078
3202	Hypothetical protein	-1.2743657
38130	Hypothetical protein	-1.2842142
105786	Hypothetical protein	-1.8152152
16630	Hypothetical protein	-1.8074216
16689	Hypothetical protein	1.4950658
21160	Hypothetical protein	-3.144422
33083	Mlh2-like protein	-1.1201041
90306	hypothetical protein	-4.5215798
31903	Hypothetical protein	-2.1003543
105642	hypothetical protein	1.2706796
16502	Hypothetical protein	-1.1714333
5657	hypothetical protein	1.0940093
11953	Coatomer alpha subunit	-1.9776048
35575	hypothetical protein	1.3882571
35090	Hypothetical protein	-1.1286587
33917	hypothetical protein	-1.0027848
38149	hypothetical protein	-1.2810748
34431	hypothetical protein	-2.0990681
32454	hypothetical protein	-2.4075854
112866	Reverse transcriptase/endonuclease, putative	1.1906261
34642	hypothetical protein	-1.8626408
15011	Hypothetical protein	-1.1753689
32460	hypothetical protein	-2.4147782
32363	Hypothetical protein	-1.125929
37931	Hypothetical protein	-1.7678753
5130	hypothetical protein	-2.1164767
11624	Hypothetical protein	-2.2750638
16841	DinF protein	-1.1070488
6147	hypothetical protein	-1.5921621

35241	hypothetical protein	-2.3063583
37871	hypothetical protein	-2.3355954
1519	hypothetical protein	-1.0590267
27806	Hypothetical protein	-1.286117
16751	Hypothetical protein	-1.4994601
39044	Hypothetical protein	-1.4859341
36095	Hypothetical protein	-1.2937277
37713	hypothetical protein	-2.3476927
37737	Hypothetical protein	-1.3040557
16693	Hypothetical protein	-1.2777013
3465	Hypothetical protein	-1.2670535
16635	hypothetical protein	-2.2588046
27623	hypothetical protein	-1.155534
33931	hypothetical protein	-2.2159347
104925	Hypothetical protein	-1.2991321
87793	Hypothetical protein	-3.8747034
113610	GlcNAc-PI synthesis protein	-1.5576588
14863	Hypothetical protein	-1.5202013
39084	Hypothetical protein	-3.0784831
16638	hypothetical protein	-1.9674039
15135	Hypothetical protein	-1.5383418
98218	hypothetical protein	-3.169746
15248	Spindle protein, putative	-1.2792175
94673	hypothetical protein	-1.0927331
8516	Hypothetical protein	-1.5694599
16247	hypothetical protein	-2.7714092
5013	Endothelin-converting enzyme 2	-1.6208002
35161	hypothetical protein	-1.113097
38903	Hypothetical protein	-1.5419471
92245	hypothetical protein	-1.0573451
3855	Hypothetical protein	-1.2329531
8454	Hypothetical protein	1.7424329
11862	hypothetical protein	-2.0230963
37700	hypothetical protein	-1.513302
19642	Hypothetical protein	-1.111936
10850	Hypothetical protein	-1.4337166
90305	hypothetical protein	-1.0280951
8650	Hypothetical protein	-1.2748754
14872	Protein 21.1	-2.6225089
28779	Hypothetical protein	-1.3983511

Appendix B

DAVID FUNCTIONAL ANNOTATION CLUSTERING OF RP7 CELL PROTEINS MODULATED IN RESPONSE TO THIAZOLIDE TREATMENT

Functional Annotation Clustering

Help and Manual

Current Gene List: Downregulated
7 DAVID IDs

Options Classification Stringency: High

Rerun using options | Create Sublist

Download File

Annotation Cluster 1	**Enrichment Score: 1.32**				**Count**	**P_Value**	**Benjamini**
GOTERM_CC_ALL	organelle lumen	RT			3	4.4E-2	1.0E0
GOTERM_CC_ALL	membrane-enclosed lumen	RT			3	4.4E-2	1.0E0
GOTERM_CC_ALL	nuclear part	RT			3	5.6E-2	1.0E0
Annotation Cluster 2	**Enrichment Score: 1.19**				**Count**	**P_Value**	**Benjamini**
GOTERM_CC_ALL	macromolecular complex	RT			6	6.2E-4	4.2E-1
GOTERM_MF_ALL	binding	RT			6	5.8E-1	1.0E0
GOTERM_BP_ALL	cellular process	RT			6	7.6E-1	1.0E0
Annotation Cluster 3	**Enrichment Score: 1.05**				**Count**	**P_Value**	**Benjamini**
GOTERM_MF_ALL	structural molecule activity	RT			3	3.6E-2	1.0E0
GOTERM_CC_ALL	cytoskeleton	RT			3	6.4E-2	1.0E0
GOTERM_CC_ALL	intracellular non-membrane-bound organelle	RT			3	1.6E-1	1.0E0
GOTERM_CC_ALL	non-membrane-bound organelle	RT			3	1.6E-1	1.0E0
Annotation Cluster 4	**Enrichment Score: 1**				**Count**	**P_Value**	**Benjamini**
SP_PIR_KEYWORDS	phosphoprotein	RT			6	6.4E-3	1.0E0
GOTERM_CC_ALL	intracellular organelle	RT			6	1.3E-1	1.0E0
GOTERM_CC_ALL	organelle	RT			6	1.3E-1	1.0E0
GOTERM_CC_ALL	intracellular part	RT			6	2.7E-1	1.0E0
GOTERM_CC_ALL	intracellular	RT			6	3.5E-1	1.0E0
Annotation Cluster 5	**Enrichment Score: 0.86**				**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	organelle organization and biogenesis	RT			3	7.4E-2	1.0E0
GOTERM_BP_ALL	cell development	RT			3	7.9E-2	1.0E0
SP_PIR_KEYWORDS	nucleotide-binding	RT			3	8.5E-2	1.0E0
GOTERM_MF_ALL	purine ribonucleotide binding	RT			3	1.4E-1	1.0E0
GOTERM_MF_ALL	ribonucleotide binding	RT			3	1.4E-1	1.0E0
GOTERM_MF_ALL	purine nucleotide binding	RT			3	1.5E-1	1.0E0
GOTERM_BP_ALL	cell differentiation	RT			3	1.5E-1	1.0E0
GOTERM_BP_ALL	cellular developmental process	RT			3	1.5E-1	1.0E0
GOTERM_MF_ALL	nucleotide binding	RT			3	1.9E-1	1.0E0
GOTERM_BP_ALL	response to stimulus	RT			3	3.4E-1	1.0E0
Annotation Cluster 6	**Enrichment Score: 0.62**				**Count**	**P_Value**	**Benjamini**
SP_PIR_KEYWORDS	Direct protein sequencing	RT			4	4.1E-2	1.0E0
GOTERM_CC_ALL	nucleus	RT			4	2.4E-1	1.0E0
GOTERM_CC_ALL	intracellular membrane-bound organelle	RT			4	5.7E-1	1.0E0
GOTERM_CC_ALL	membrane-bound organelle	RT			4	5.7E-1	1.0E0
Annotation Cluster 7	**Enrichment Score: 0.53**				**Count**	**P_Value**	**Benjamini**
SP_PIR_KEYWORDS	disease mutation	RT			3	7.6E-2	1.0E0
GOTERM_CC_ALL	cytoplasmic part	RT			3	4.3E-1	1.0E0
UP_SEQ_FEATURE	sequence variant	RT			3	8.1E-1	1.0E0
Annotation Cluster 8	**Enrichment Score: 0.34**				**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	cellular metabolic process	RT			5	4.2E-1	1.0E0
GOTERM_BP_ALL	primary metabolic process	RT			5	4.2E-1	1.0E0
GOTERM_BP_ALL	metabolic process	RT			5	5.4E-1	1.0E0
Annotation Cluster 9	**Enrichment Score: 0.22**				**Count**	**P_Value**	**Benjamini**

	Annotation Cluster 1	Enrichment Score: 1.32	G		Count	P_Value	Benjamini
☐	GOTERM_BP_ALL	regulation of cellular process	RT		3	5.5E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of biological process	RT		3	6.0E-1	1.0E0
☐	GOTERM_BP_ALL	biological regulation	RT		3	6.6E-1	1.0E0

19 terms were not clustered.

Functional Annotation Clustering

Help and Manual

Current Gene List: Upregulated
32 DAVID IDs

Options **Classification Stringency** High

Rerun using options | Create Sublist

Download File

Annotation Cluster 1	**Enrichment Score: 4.13**			**Count**	**P_Value**	**Benjamini**
SP_PIR_KEYWORDS	molecular chaperone	RT		5	1.4E-8	7.5E-6
INTERPRO	Heat shock protein Hsp70	RT		4	2.6E-6	1.5E-2
INTERPRO	Heat shock protein 70	RT		4	4.5E-6	1.3E-2
PIR_SUPERFAMILY	PIRSF002581:chaperone HSP70	RT		4	4.9E-6	1.5E-2
SP_PIR_KEYWORDS	chaperone	RT		5	8.9E-5	1.6E-2
SP_PIR_KEYWORDS	stress response	RT		4	2.5E-4	3.3E-2
GOTERM_BP_ALL	response to protein stimulus	RT		4	6.5E-4	6.8E-1
GOTERM_BP_ALL	response to unfolded protein	RT		4	6.5E-4	6.8E-1
SP_PIR_KEYWORDS	ATP	RT		5	8.7E-4	9.8E-2
GOTERM_BP_ALL	response to biotic stimulus	RT		5	2.4E-3	7.5E-1
GOTERM_BP_ALL	response to chemical stimulus	RT		5	2.3E-2	9.9E-1
Annotation Cluster 2	**Enrichment Score: 3.42**			**Count**	**P_Value**	**Benjamini**
GOTERM_CC_ALL	intracellular membrane-bound organelle	RT		22	2.4E-4	5.0E-2
GOTERM_CC_ALL	membrane-bound organelle	RT		22	2.4E-4	4.1E-2
GOTERM_CC_ALL	intracellular organelle	RT		23	5.9E-4	4.6E-2
GOTERM_CC_ALL	organelle	RT		23	6.0E-4	4.2E-2
Annotation Cluster 3	**Enrichment Score: 3**			**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	anti-apoptosis	RT		5	2.4E-4	4.6E-1
GOTERM_BP_ALL	negative regulation of apoptosis	RT		5	7.6E-4	5.5E-1
GOTERM_BP_ALL	negative regulation of programmed cell death	RT		5	7.9E-4	5.0E-1
GOTERM_BP_ALL	regulation of apoptosis	RT		6	2.6E-3	6.5E-1
GOTERM_BP_ALL	regulation of programmed cell death	RT		6	2.7E-3	6.4E-1
Annotation Cluster 4	**Enrichment Score: 2.75**			**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	cellular protein metabolic process	RT		15	1.4E-3	6.5E-1
GOTERM_BP_ALL	cellular macromolecule metabolic process	RT		15	1.6E-3	6.6E-1
GOTERM_BP_ALL	protein metabolic process	RT		15	2.5E-3	7.0E-1
Annotation Cluster 5	**Enrichment Score: 2.2**			**Count**	**P_Value**	**Benjamini**
GOTERM_CC_ALL	intracellular part	RT		26	8.0E-5	2.3E-2
GOTERM_CC_ALL	intracellular	RT		26	3.5E-4	4.9E-2
GOTERM_CC_ALL	cell part	RT		27	2.4E-1	1.0E0
GOTERM_CC_ALL	cell	RT		27	2.4E-1	1.0E0
Annotation Cluster 6	**Enrichment Score: 2.09**			**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	apoptosis	RT		7	2.5E-3	7.3E-1
GOTERM_BP_ALL	programmed cell death	RT		7	2.6E-3	6.8E-1
GOTERM_BP_ALL	regulation of apoptosis	RT		6	2.6E-3	6.5E-1
GOTERM_BP_ALL	regulation of programmed cell death	RT		6	2.7E-3	6.4E-1
GOTERM_BP_ALL	death	RT		7	3.4E-3	7.0E-1

					Count	P_Value	Benjamini
Annotation Cluster 1	Enrichment Score: 4.13						
GOTERM_BP_ALL	cell death	RT			7	3.4E-3	7.0E-1
GOTERM_BP_ALL	cell development	RT			7	2.2E-2	9.9E-1
GOTERM_BP_ALL	cellular developmental process	RT			7	1.1E-1	1.0E0
GOTERM_BP_ALL	cell differentiation	RT			7	1.1E-1	1.0E0
Annotation Cluster 7	Enrichment Score: 1.82				Count	P_Value	Benjamini
GOTERM_BP_ALL	metabolic process	RT			24	1.0E-2	9.5E-1
GOTERM_BP_ALL	cellular metabolic process	RT			22	1.8E-2	9.9E-1
GOTERM_BP_ALL	primary metabolic process	RT			22	1.9E-2	9.9E-1
Annotation Cluster 8	Enrichment Score: 1.8				Count	P_Value	Benjamini
GOTERM_BP_ALL	intracellular transport	RT			6	8.6E-3	9.3E-1
GOTERM_BP_ALL	establishment of cellular localization	RT			6	2.0E-2	9.9E-1
GOTERM_BP_ALL	cellular localization	RT			6	2.2E-2	9.9E-1
Annotation Cluster 9	Enrichment Score: 1.37				Count	P_Value	Benjamini
SP_PIR_KEYWORDS	nucleotide-binding	RT			8	1.3E-2	7.6E-1
SP_PIR_KEYWORDS	atp-binding	RT			7	1.4E-2	7.2E-1
GOTERM_MF_ALL	ribonucleotide binding	RT			8	5.6E-2	1.0E0
GOTERM_MF_ALL	purine ribonucleotide binding	RT			8	5.6E-2	1.0E0
GOTERM_MF_ALL	ATP binding	RT			7	5.8E-2	1.0E0
GOTERM_MF_ALL	adenyl ribonucleotide binding	RT			7	6.1E-2	1.0E0
GOTERM_MF_ALL	purine nucleotide binding	RT			8	6.7E-2	1.0E0
GOTERM_MF_ALL	adenyl nucleotide binding	RT			7	7.4E-2	1.0E0
Annotation Cluster 10	Enrichment Score: 1.08				Count	P_Value	Benjamini
GOTERM_BP_ALL	establishment of localization	RT			10	5.4E-2	1.0E0
GOTERM_BP_ALL	transport	RT			9	1.0E-1	1.0E0
GOTERM_BP_ALL	localization	RT			10	1.0E-1	1.0E0
Annotation Cluster 11	Enrichment Score: 1.04				Count	P_Value	Benjamini
SMART	RRM	RT			3	2.6E-2	1.0E0
INTERPRO	RNA recognition motif, RNP-1	RT			3	5.4E-2	1.0E0
INTERPRO	Nucleotide-binding, alpha-beta plait	RT			3	5.6E-2	1.0E0
SP_PIR_KEYWORDS	rna-binding	RT			3	2.0E-1	1.0E0
GOTERM_MF_ALL	RNA binding	RT			3	4.0E-1	1.0E0
Annotation Cluster 12	Enrichment Score: 0.93				Count	P_Value	Benjamini
GOTERM_MF_ALL	nucleoside-triphosphatase activity	RT			4	1.1E-1	1.0E0
GOTERM_MF_ALL	pyrophosphatase activity	RT			4	1.2E-1	1.0E0
GOTERM_MF_ALL	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	RT			4	1.2E-1	1.0E0
GOTERM_MF_ALL	hydrolase activity, acting on acid anhydrides	RT			4	1.2E-1	1.0E0
Annotation Cluster 13	Enrichment Score: 0.92				Count	P_Value	Benjamini
GOTERM_BP_ALL	intracellular protein transport	RT			4	3.7E-2	1.0E0
GOTERM_BP_ALL	protein transport	RT			4	1.3E-1	1.0E0
GOTERM_BP_ALL	establishment of protein localization	RT			4	1.5E-1	1.0E0
GOTERM_BP_ALL	protein localization	RT			4	1.7E-1	1.0E0
GOTERM_BP_ALL	macromolecule localization	RT			4	2.0E-1	1.0E0
Annotation Cluster 14	Enrichment Score: 0.69				Count	P_Value	Benjamini
UP_SEQ_FEATURE	transit peptide: Mitochondrion	RT			3	1.5E-1	1.0E0
SP_PIR_KEYWORDS	transit peptide	RT			3	1.6E-1	1.0E0

	Annotation Cluster 1	Enrichment Score: 4.13			Count	P_Value	Benjamini
☐	SP_PIR_KEYWORDS	Mitochondrion	RT		3	3.6E-1	1.0E0
	Annotation Cluster 15	**Enrichment Score: 0.34**			**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_ALL	nervous system development	RT		4	1.7E-1	1.0E0
☐	GOTERM_BP_ALL	system development	RT		5	4.0E-1	1.0E0
☐	GOTERM_BP_ALL	anatomical structure development	RT		5	5.7E-1	1.0E0
☐	GOTERM_BP_ALL	multicellular organismal development	RT		5	6.4E-1	1.0E0
☐	GOTERM_BP_ALL	multicellular organismal process	RT		6	8.5E-1	1.0E0
	Annotation Cluster 16	**Enrichment Score: 0.28**			**Count**	**P_Value**	**Benjamini**
☐	GOTERM_MF_ALL	ion binding	RT		10	4.1E-1	1.0E0
☐	GOTERM_MF_ALL	metal ion binding	RT		9	5.5E-1	1.0E0
☐	GOTERM_MF_ALL	cation binding	RT		8	6.2E-1	1.0E0
	Annotation Cluster 17	**Enrichment Score: 0.06**			**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_ALL	regulation of metabolic process	RT		5	8.2E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of gene expression	RT		4	9.0E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of cellular metabolic process	RT		4	9.2E-1	1.0E0
	Annotation Cluster 18	**Enrichment Score: 0.02**			**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_ALL	regulation of transcription	RT		3	9.6E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		3	9.7E-1	1.0E0
☐	GOTERM_BP_ALL	transcription	RT		3	9.7E-1	1.0E0

119 terms were not clustered.

Please cite *Nature Protocols 2009; 4(1):44 & Genome Biology 2003; 4(5):P3* within any publication that makes use of any methods inspired by **DAVID**.

Functional Annotation Clustering

Help and Manual

Current Gene List: All proteins
37 DAVID IDs

Options **Classification Stringency** High

Rerun using options | Create Sublist

Download File

						Count	P_Value	Benjamini
Annotation Cluster 1	**Enrichment Score: 4.16**							
	SP_PIR_KEYWORDS	molecular chaperone	RT			5	2.6E-8	7.0E-6
	INTERPRO	Heat shock protein Hsp70	RT			4	4.1E-6	2.4E-2
	PIR_SUPERFAMILY	PIRSF002581:chaperone HSP70	RT			4	7.0E-6	2.2E-2
	INTERPRO	Heat shock protein 70	RT			4	7.1E-6	2.1E-2
	SP_PIR_KEYWORDS	chaperone	RT			5	1.6E-4	2.4E-2
	SP_PIR_KEYWORDS	stress response	RT			4	3.9E-4	5.1E-2
	GOTERM_BP_ALL	response to unfolded protein	RT			4	1.1E-3	6.7E-1
	GOTERM_BP_ALL	response to protein stimulus	RT			4	1.1E-3	6.7E-1
	SP_PIR_KEYWORDS	ATP	RT			5	1.5E-3	1.5E-1
	GOTERM_BP_ALL	response to biotic stimulus	RT			5	4.3E-3	7.4E-1
Annotation Cluster 2	**Enrichment Score: 3.01**							
	GOTERM_BP_ALL	anti-apoptosis	RT			5	4.5E-4	6.9E-1
	GOTERM_BP_ALL	negative regulation of apoptosis	RT			5	1.4E-3	5.7E-1
	GOTERM_BP_ALL	negative regulation of programmed cell death	RT			5	1.5E-3	5.4E-1
Annotation Cluster 3	**Enrichment Score: 2.7**							
	GOTERM_CC_ALL	intracellular part	RT			31	9.0E-6	2.6E-3
	GOTERM_CC_ALL	intracellular	RT			31	5.3E-5	4.1E-3
	GOTERM_CC_ALL	cell part	RT			32	1.9E-1	9.9E-1
	GOTERM_CC_ALL	cell	RT			32	1.9E-1	9.9E-1
Annotation Cluster 4	**Enrichment Score: 2.65**							
	GOTERM_CC_ALL	cytoskeleton	RT			9	1.2E-3	6.5E-2
	GOTERM_CC_ALL	intracellular non-membrane-bound organelle	RT			11	3.0E-3	1.3E-1
	GOTERM_CC_ALL	non-membrane-bound organelle	RT			11	3.0E-3	1.3E-1
Annotation Cluster 5	**Enrichment Score: 2.53**							
	GOTERM_BP_ALL	regulation of apoptosis	RT			7	8.9E-4	7.9E-1
	GOTERM_BP_ALL	regulation of programmed cell death	RT			7	9.4E-4	7.1E-1
	GOTERM_BP_ALL	apoptosis	RT			8	1.2E-3	5.8E-1
	GOTERM_BP_ALL	programmed cell death	RT			8	1.2E-3	5.6E-1
	GOTERM_BP_ALL	cell death	RT			8	1.7E-3	5.5E-1
	GOTERM_BP_ALL	death	RT			8	1.7E-3	5.5E-1
	GOTERM_BP_ALL	cell development	RT			9	4.0E-3	7.4E-1
	GOTERM_BP_ALL	cellular developmental process	RT			9	3.7E-2	1.0E0
	GOTERM_BP_ALL	cell differentiation	RT			9	3.7E-2	1.0E0
Annotation Cluster 6	**Enrichment Score: 2.4**							
	GOTERM_BP_ALL	cellular protein metabolic process	RT			16	3.1E-3	7.2E-1
	GOTERM_BP_ALL	cellular macromolecule metabolic process	RT			16	3.7E-3	7.5E-1
	GOTERM_BP_ALL	protein metabolic process	RT			16	5.7E-3	8.1E-1

Annotation Cluster 1	Enrichment Score: 4.16				Count	P_Value	Benjamini
Annotation Cluster 7	Enrichment Score: 2.12				Count	P_Value	Benjamini
GOTERM_BP_ALL	intracellular transport	RT			7	3.7E-3	7.3E-1
GOTERM_BP_ALL	establishment of cellular localization	RT			7	1.0E-2	9.1E-1
GOTERM_BP_ALL	cellular localization	RT			7	1.2E-2	9.2E-1
Annotation Cluster 8	Enrichment Score: 1.74				Count	P_Value	Benjamini
SP_PIR_KEYWORDS	nucleotide-binding	RT			10	2.5E-3	2.1E-1
SP_PIR_KEYWORDS	atp-binding	RT			8	8.1E-3	4.8E-1
GOTERM_MF_ALL	purine ribonucleotide binding	RT			10	1.7E-2	1.0E0
GOTERM_MF_ALL	ribonucleotide binding	RT			10	1.7E-2	1.0E0
GOTERM_MF_ALL	purine nucleotide binding	RT			10	2.2E-2	1.0E0
GOTERM_MF_ALL	ATP binding	RT			8	4.1E-2	1.0E0
GOTERM_MF_ALL	adenyl ribonucleotide binding	RT			8	4.4E-2	1.0E0
GOTERM_MF_ALL	adenyl nucleotide binding	RT			8	5.5E-2	1.0E0
Annotation Cluster 9	Enrichment Score: 1.7				Count	P_Value	Benjamini
GOTERM_BP_ALL	metabolic process	RT			27	1.7E-2	9.7E-1
GOTERM_BP_ALL	cellular metabolic process	RT			25	2.1E-2	9.8E-1
GOTERM_BP_ALL	primary metabolic process	RT			25	2.2E-2	9.8E-1
Annotation Cluster 10	Enrichment Score: 1.31				Count	P_Value	Benjamini
GOTERM_BP_ALL	macromolecule catabolic process	RT			5	1.4E-2	9.5E-1
GOTERM_BP_ALL	cellular macromolecule catabolic process	RT			4	3.9E-2	1.0E0
GOTERM_BP_ALL	catabolic process	RT			5	7.4E-2	1.0E0
GOTERM_BP_ALL	cellular catabolic process	RT			4	1.3E-1	1.0E0
Annotation Cluster 11	Enrichment Score: 1.3				Count	P_Value	Benjamini
GOTERM_MF_ALL	nucleoside-triphosphatase activity	RT			5	4.5E-2	1.0E0
GOTERM_MF_ALL	pyrophosphatase activity	RT			5	5.1E-2	1.0E0
GOTERM_MF_ALL	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	RT			5	5.2E-2	1.0E0
GOTERM_MF_ALL	hydrolase activity, acting on acid anhydrides	RT			5	5.4E-2	1.0E0
Annotation Cluster 12	Enrichment Score: 1.25				Count	P_Value	Benjamini
SMART	RRM	RT			3	3.5E-2	1.0E0
INTERPRO	RNA recognition motif, RNP-1	RT			3	7.0E-2	1.0E0
INTERPRO	Nucleotide-binding, alpha-beta plait	RT			3	7.3E-2	1.0E0
Annotation Cluster 13	Enrichment Score: 1.13				Count	P_Value	Benjamini
GOTERM_BP_ALL	localization	RT			12	6.1E-2	1.0E0
GOTERM_BP_ALL	establishment of localization	RT			11	6.3E-2	1.0E0
GOTERM_BP_ALL	transport	RT			10	1.1E-1	1.0E0
Annotation Cluster 14	Enrichment Score: 0.77				Count	P_Value	Benjamini
GOTERM_BP_ALL	intracellular protein transport	RT			4	5.5E-2	1.0E0
GOTERM_BP_ALL	protein transport	RT			4	1.9E-1	1.0E0
GOTERM_BP_ALL	establishment of protein localization	RT			4	2.1E-1	1.0E0
GOTERM_BP_ALL	protein localization	RT			4	2.4E-1	1.0E0
GOTERM_BP_ALL	macromolecule localization	RT			4	2.7E-1	1.0E0
Annotation Cluster 15	Enrichment Score: 0.59				Count	P_Value	Benjamini
UP_SEQ_FEATURE	transit peptide: Mitochondrion	RT			3	1.9E-1	1.0E0
SP_PIR_KEYWORDS	transit peptide	RT			3	2.0E-1	1.0E0
SP_PIR_KEYWORDS	Mitochondrion	RT			3	4.4E-1	1.0E0
Annotation Cluster 16	Enrichment Score: 0.52				Count	P_Value	Benjamini

	Annotation Cluster 1	Enrichment Score: 4.16	G		Count	P_Value	Benjamini
☐	GOTERM_BP_ALL	system development	RT		7	1.7E-1	1.0E0
☐	GOTERM_BP_ALL	anatomical structure development	RT		8	1.7E-1	1.0E0
☐	GOTERM_BP_ALL	multicellular organismal development	RT		7	3.9E-1	1.0E0
☐	GOTERM_BP_ALL	multicellular organismal process	RT		8	7.3E-1	1.0E0
	Annotation Cluster 17	**Enrichment Score: 0.05**	**G**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_ALL	regulation of metabolic process	RT		6	7.9E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of gene expression	RT		5	8.7E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of cellular metabolic process	RT		5	8.9E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of transcription	RT		4	9.3E-1	1.0E0
☐	GOTERM_BP_ALL	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		4	9.4E-1	1.0E0
☐	GOTERM_BP_ALL	transcription	RT		4	9.5E-1	1.0E0
	Annotation Cluster 18	**Enrichment Score: 0.01**	**G**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_ALL	regulation of transcription, DNA-dependent	RT		3	9.8E-1	1.0E0
☐	GOTERM_BP_ALL	transcription, DNA-dependent	RT		3	9.8E-1	1.0E0
☐	GOTERM_BP_ALL	RNA biosynthetic process	RT		3	9.8E-1	1.0E0

143 terms were not clustered.
